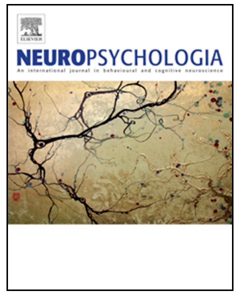

Theory of mind and decision science: Towards a typology of tasks and computational models

Tessa Rusch, Saurabh Steixner-Kumar, Prashant Doshi, Michael Spezio, Jan Gläscher

PII: S0028-3932(20)30159-7

DOI: https://doi.org/10.1016/j.neuropsychologia.2020.107488

Reference: NSY 107488

To appear in: *Neuropsychologia*

Received Date: 30 September 2019

Revised Date: 27 April 2020

Accepted Date: 4 May 2020

CRediT author statement

Tessa Rusch: Conceptualization, Writing - Original Draft, Writing - Review & Editing, Visualization. **Saurabh Steixner-Kumar**: Conceptualization, Writing - Review & Editing. **Prashant Doshi**: Conceptualization, Writing - Review & Editing. **Michael Spezio:** Conceptualization, Writing - Review & Editing, Funding acquisition. **Jan Gläscher**: Conceptualization, Writing - Original Draft, Writing - Review & Editing, Funding acquisition.

Theory of Mind and Decision Science: Towards a Typology of Tasks and Computational Models

Tessa Rusch[1,2+], Saurabh Steixner-Kumar[1], Prashant Doshi[3], Michael Spezio*[,+ 1,4] & Jan Gläscher* [1]

* shared senior author

[+] corresponding authors: Tessa Rusch, tessa.rusch@gmail.com; Michael Spezio, mspezio@scrippscollege.edu

[1] Institute of Systems Neuroscience, University Medical Center Hamburg-Eppendorf, Martinistr. 52, 20251 Hamburg

[2] Division of the Humanities and Social Sciences, California Institute of Technology, 1200 E. California Blvd., Pasadena, CA 91125

[3] Department of Computer Science, University of Georgia, 539 Boyd GSRC, Athens, GA 30602

[4] Psychology, Neuroscience, and Data Science, Scripps College, 1030 N Columbia Ave, Claremont, CA 91711

Keywords

Theory of Mind; computational modeling; decision making; interactivity; uncertainty

Abstract

The ability to form a Theory of Mind (ToM), i.e., to theorize about others' mental states to explain and predict behavior in relation to attributed intentional states, constitutes a hallmark of human cognition. These abilities are multi-faceted and include a variety of different cognitive sub-functions. Here, we focus on decision processes in social contexts and review a number of experimental and computational modeling approaches in this field. We provide an overview of experimental accounts and formal computational models with respect to two dimensions: interactivity and uncertainty. Thereby, we aim at capturing the nuances of ToM functions in the context of social decision processes. We suggest there to be an increase in ToM engagement and multiplexing as social cognitive decision-making tasks become more interactive and uncertain. We propose that representing others as intentional and goal directed agents who perform consequential actions is elicited only at the edges of these two dimensions. Further, we argue that computational models of valuation and beliefs follow these dimensions to best allow researchers to effectively model sophisticated ToM-processes. Finally, we relate this typology to neuroimaging findings in neurotypical (NT) humans, studies of persons with autism spectrum (AS), and studies of nonhuman primates.

1. Introduction

Humans are distinctly skilled at sophisticated social interactions. To successfully engage in social exchanges, they rely on "Theory of Mind" (ToM). ToM is a concept defined by Premack and Woodruff (1978) in the highly influential article "Does the Chimpanzee have a theory of mind?" as "an individual imputing mental states [like beliefs, desires and intentions] to himself and others […] to make predictions, specifically about the behavior of other organisms". In their paper, Premack and Woodruff stressed that ToM need not be accurate for it to be present (i.e., false inferences often do result from its presence and not exclusively due to its absence). Further, they differentiated between ToM for motivation (i.e., another organism's valuation, intention, purpose, goal) and ToM for knowledge (i.e., another organism's belief states or learned schemas/scripts). Along with this comprehensive definition, Premack and Woodruff proposed an initial task to probe ToM capacities in nonhuman primates: they presented short videos to a chimpanzee named Sarah of a human struggling with simple tasks. Subsequently, Sarah saw photographs of various items, including the one solving the actor's problem. Sarah's ability to select the correct photograph served as evidence of her being capable of recognizing the problem (i.e., representing the state of affairs, as well as the actor's purpose in the scene, so his intentions and goals). This highlighted that having a Theory of Mind requires a representation of the state of affairs and a representation of an individual's purposeful and motivational relationship to that state, i.e., the individual's beliefs and values/goals, respectively, in the situation (Wimmer and Perner, 1983). ToM is thus not a unitary process. ToM is instead a category that includes at least two differentiable social cognitive processes capable of representing the first order beliefs and first order values attributed to others, along with processes for sharing and integrating these representations.

Since this initial empirical investigation into ToM in nonhuman primates, experimental approaches probing and characterizing ToM capacities have been introduced by psychological and behavioral economics research (Houser and McCabe, 2014; Kovacs et al., 2010; Schurz et al., 2014; Wimmer and Perner, 1983). Neural networks implicated in ToM were successfully identified using standard neuroimaging methods (Gallagher and Frith, 2003; Schurz et al., 2014; Siegal and Varley, 2002). Further, analyses of neural signals increasingly drew on quantitative descriptions of covert cognitive processes constituting ToM via computational models of behavior (Charpentier & O'Doherty, 2018; Hampton, Bossaerts, & O'Doherty, 2008; Hill et al., 2017; Xiao, Geng, Riggins, Chen, & Redcay, 2019; Yoshida et al., 2010). Despite the vast success of these approaches, a coherent picture of what ToM is, how humans and other species engage in it, and which neural mechanisms constitute it, is missing (Emery and Clayton, 2009; Schaafsma et al., 2015). We argue that in part this is due to graded differences in the cognitive processes elicited by various ToM tasks. More specifically, we propose that the extent to which they require an intentional representation of other individuals and the degree of integration between such representations of others and one's own reference frame is highly variable.

Premack and Woodruff's conceptual differentiation of ToM's knowledge and motivational processes has been followed by other investigators, distinguishing between so-called cognitive and affective ToM (Baron-Cohen, 1988; Kalbe et al., 2010; Mitchell & Phillips, 2015). In these accounts, "cognitive ToM" is primarily focused on explicit perspective-taking and strategic reasoning about another person's beliefs, generating causal inferences and predictions about the other's behavior. The term "affective ToM" in most investigations is restricted to cognitive processes of inference about the emotions of others, such as empathy, emotion recognition and emotion simulation, and typically does not emphasize goal states or valuations of possible actions. Both cognitive and affective ToM processes have been investigated in great detail. Examining both lines of research at once would go beyond the scope of a single article. Therefore, in this current review, we exclusively focus on perspective taking and valuational and motivational ToM processes during decision problems. Such processes can be considered affective just as they are cognitive. However, in this paper, we do not explicitly consider the processes more typically denoted as affective, such as empathy for emotional states. Instead, we examine how decision tasks aimed at ToM likely differ with respect to the cognitive functions they elicit. We present a typology of experimental approaches and cognitive computational models along two primary dimensions: *interactivity* and *uncertainty*. We propose that this typology can help to interpret existing findings on the behavioral and neural levels and can aid task design in future studies. Specifically, we suggest that tasks which combine higher levels of both uncertainty and interactivity facilitate investigations of and potentially provide greater insight into high-level ToM.

1.1. The functional relevance of interactivity and uncertainty in ToM tasks

In the presented typology of ToM tasks, uncertainty refers to either the risk or ambiguity characterizing associations between actions, states, state transitions, and outcomes (Hsu et al., 2005). Additionally, although formally not covered by uncertainty, we discuss the availability and accessibility of information in the context of uncertainty. We propose that under uncertainty and unequal distribution of information between agents, others' intentional states likely become highly relevant and distinguishable from one's own intentional states. However, task risks or ambiguities must not be so great as to be simply random chance or there will be little incentive for any learning and therefore little incentive for tracking others' intentional states. Uncertainty occurs both for environmental (e.g., state, reward) and social (i.e., agent) variables, and both are relevant to the roles that ToM plays in choice and behavior. Environmental uncertainties may arise when joint action-outcome associations or state transitions are probabilistic, and their dynamical changes are unknown. Social uncertainty refers to the uncertainty about the other agents' actions, because their preferences, goals, beliefs, abilities to track the environmental variables, rationality or stochasticity, etc., are unknown.

Interactivity (Byom and Mutlu, 2013; Jording et al., 2019) in our proposed typology of social cognition tasks refers to a combination of the social distance or face-to-face context (e.g., still photos, recorded video, live video, interactive live video, interactive in person; Spezio, Huang, Castelli, & Adolphs, 2007), the personal relevance, the task-dependent consequences of a social cognition task (Bublatzky et al., 2017), and the level of involvement of multiple agents (Norris et al., 2014). Interactivity is a dimension of socially oriented tasks that ranges from purely passive spectatorial observation to full consequential interaction. Thereby, interactivity determines the behavioral relevance of ToM. Behavioral relevance is understood as the relative importance of making predictions of others' behavior from their frames of reference, using those predictions to plan one's own (re)actions, and so integrating predictions from ToM into one's own perspective.

We begin by summarizing a range of relevant ToM tasks from psychology, economics and decision neuroscience, and characterize the different experimental approaches based on the two proposed dimensions. We suggest that divergent knowledge about the environment due to unshared information and asymmetric environmental uncertainties motivate the representations of others' belief states while social uncertainties elicit representations of others' motivational states. If all information is equally accessible to all agents involved in a task, participants observing or interacting with other agents have no need for ToM beyond positing that another rational, competent agent wants to succeed in the task and has beliefs that correctly conform to the task contingencies. As risk or ambiguity increases and different information about the environment become available to the participants and the agents they observe or interact with, participants must distinguish their own assessment of the environment from the other agents' assessments (i.e., beliefs about the states, about the state transitions, or about the reward outcomes). As the other agents' motivations, intentions and reasoning processes become unclear an increased demand to represent motivational states is created.

Second, tasks are characterized with respect to the type of interactivity they include. We argue that the degree of interactivity and active engagement influences the need to take others' perspectives and influences the level of interaction of such representations with self-referential processes. The distinction between self- and other-referential processes in the realm of social decision making has proven very useful for the functional relevance of different brain networks relevant during social learning (Joiner et al., 2017; Qu et al., 2017). We follow this differentiation and discuss self- and other-referential cognitive processes and their interaction depending on the varying levels of interactivity in experimental paradigms. We propose that a task, where participants passively observe others' actions in a context that entails no requirement for any response or judgement nor any consequences for the observer, requires less ToM and less self-referential processing than a task where participants are personally involved with another agent with gains and losses dependent upon the decisions made by both. The function of ToM in the latter case would be to enhance the accurate predictions of the other's actions and so to improve successful coordination or competition. Thus, social tasks in which multiple agents interact cooperatively or competitively in real time with real consequences could foster higher levels of ToM than less interactive tasks where little or nothing is at stake. In synchronous, interactive, consequential tasks, participants would be expected to represent another agent's representation of themselves (second-level ToM) or even go farther in tasks requiring complex synchronous interaction to achieve task-relevant goals (Doshi, Qu, Goodie, & Young, 2012, Doshi, Qu, & Goodie, 2014).

In the second section of this review, we examine different computational models that have been used to quantify the cognitive processes individuals engage in when solving such tasks and characterize models with respect to the aforementioned dimensions. Lastly, we interpret neural findings in neurotypical (NT) humans.

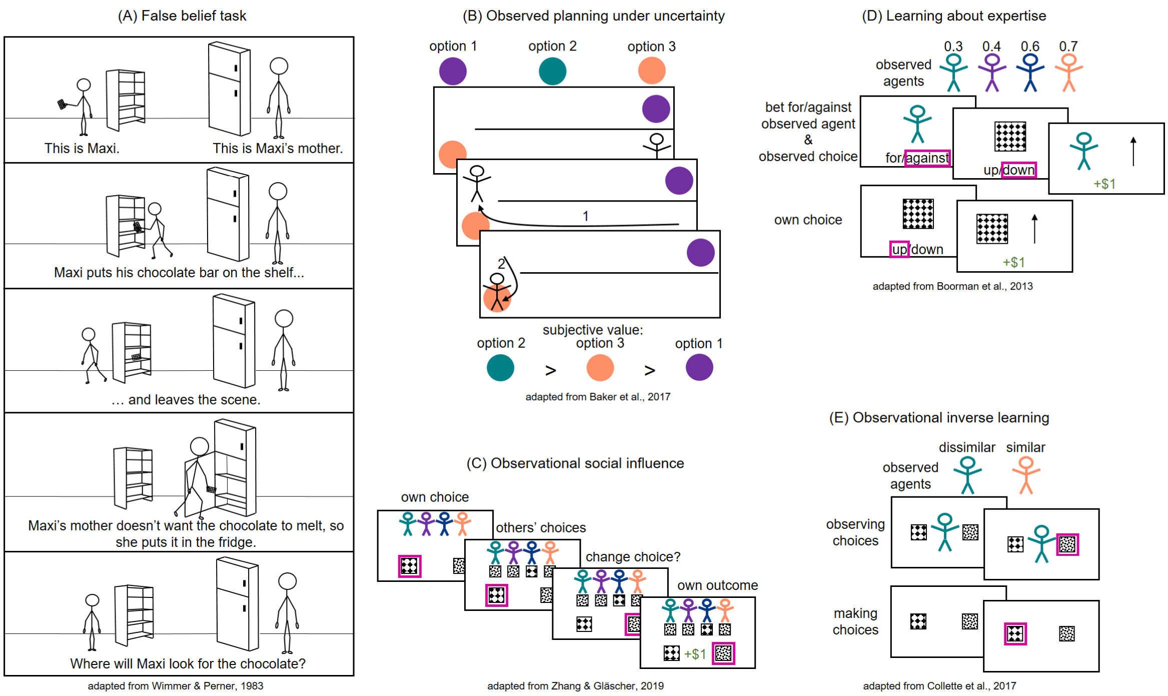

Figure 1| Observational ToM tasks comprising environmental uncertainty.
(A) In the false belief task, a simple social scenario is presented to a participant. After observing a social scene that comprises a change in the physical environment that the observed agent is unaware of (inducing a false belief), participants have to predict that observed agent's behavior. To successfully do so, participants have to differentiate their own representation of the environment from the observed agent's (physically inaccurate) perspective. **(B)** The trajectory of an agent starting from the bottom right corner of a simple maze-like environment is presented to participants. The observed agent's perceptual abilities are limited by occluding walls preventing them from overseeing the entire scene. The participant takes a bird's eye view. Based on the path that the agent takes (here indicated by arrows) participants are asked to indicate the agent's subjective preferences over available goal states (here: purple, green and orange). **(C)** In a group decision game, participants need to learn the value of two dynamically changing probabilistically rewarded choice options. After making their own choice, the selections made by other players who are learning about the same choice options are revealed and participants are allowed to adjust their choice if desired. Finally, feedback about the reward outcome associated with the chosen option is presented. **(D)** Participants observe four different co-players with varying expertise in a probabilistic value learning task. First, they choose between betting for or against these agents' success. Second, they see the observed agent's choice (their predictions about whether the presented asset would increase or decrease in value). Last, they receive feedback about whether their bet was correct or not by either winning or losing money, allowing inference about the others' expertise and the value of assets. **(E)** Participants observe the actions of different agents whose preferences they learned in a pre-test training period. The outcome reward associated with those actions is not revealed to the observer. To infer the underlying reward distribution, participants need to represent the observed agents' learning processes and interpret the observed agents' actions in light of their preferences.

2. Tasks

2.1. Observation under divergent knowledge and environmental uncertainty

2.1.1. False belief reasoning and perspective taking

One of the most prominent tasks in ToM research is the so-called false belief task (Figure 1A) first formulated by Wimmer and Perner (1983). It is a short story where the character Maxi puts a chocolate bar on a shelf and leaves the scene. In Maxi's absence, his mother changes the state of affairs by moving the chocolate bar to a different location. Upon Maxi's return, the observing participant is asked where Maxi would search for his chocolate. The key feature in this task is the change in the state of affairs and Maxi's ignorance of that change, i.e., Maxi's false belief. Unknown to Maxi, the contingencies of the environment he acts in changed. This means, that his limited knowledge about the environment leads him to a false belief. In contrast, the observing participant has perfect knowledge of the environment. To correctly predict Maxi's behavior, observing participants need to differentiate their own correct belief about the situation from their representation of Maxi's false belief and respond based on their representation of Maxi's mental state. Variants of the false belief task have been deployed to assess the development of ToM abilities in children, differences in individuals with Autism Spectrum (AS), and nonhuman species' abilities to reason about others (e.g. Baillargeon, Scott, & He, 2010; Baron-Cohen, Leslie, & Frith, 1985; Bora, Yucel, & Pantelis, 2009; Call & Tomasello, 1999; Dufour et al., 2013; Saxe & Kanwisher, 2003; Wimmer & Perner, 1983). Common to most of these variations is the use of social scenes that require judgment about a false belief scenario. Yet, depending on task specifics, findings about when healthy children develop the ability to theorize about other minds differ. When explicitly asked, children typically answer questions about an agent's false belief correctly from around four years on (Wimmer and Perner, 1983). However, 13-month-old infants show correct anticipatory viewing behavior in such tasks (Surian et al., 2007) potentially suggesting an earlier onset of false belief understanding (Baillargeon et al., 2010).

Following a similar general idea as false belief reasoning, Baker and colleagues (Baker et al., 2017) introduced a perspective taking scenario which required putting oneself in someone else's shoes and seeing the world from their eyes. They used maze-like spatial layouts, an environment well suited for the application of formal decision models, to examine inferential processes about an observed agent's beliefs and desires: an observed agent with unknown preferences is placed in an environment containing different choice options with varying subjective value to the agent (Figure 1B). At any given trial, only a subset of options is available in the environment. Additionally, occluding walls prevent the agent from overseeing the entire space. The agent has to move around to explore what options are currently available and then choose the option that is most valuable to him. Participants observe the agent while taking a bird's eye view. As in the false belief task, participants are fully informed about the environmental properties, but the observed agent is uninformed about the availability of goal states. That means, participants and observed agents have asymmetric knowledge about the environment, and the observed agent is faced with uncertainty about the availability of goal states. Additionally, the observed agent's preferences regarding choice options are unknown to participants. Figure 1B shows an exemplary situation. Two out of three possible choice options (here indicated by purple, green and orange) of varying subjective value to the observed agent are available. From the initial position, the observed agent can only see the orange option. The agent first moves around the occluding wall but then turns around and returns to the orange option. When asked to rate the agent's preferences based on this behavior, participants indicate that green is most valuable to the agent followed by orange and rate purple as least valuable. In search of the most preferred option, green, the agent moved around the wall but returned to the second-best option, orange, when seeing that only purple, the least favorable option, was available. These judgments indicate that participants infer the agent's valuations based on the agent's perceptual experience, meaning based on the choice options he can and cannot see, and attribute preferences to explain the agent's movements.

In false belief and observational perspective taking tasks, information about the environment is distributed unequally between participants and observed agents. Fully informed participants watch and predict agents acting in environments accurately known to the participants but not the agents. This task element induces a divergence between participants' own knowledge about the environment and the observed agents' knowledge about the environment. This mismatch is a key factor for triggering reasoning about another person's knowledge state. If information about the environment is equally accessible to all participants and agents, there is little reason to take others' perspectives as it provides little or no additional information about the shared state. However, by inducing differences between one's own and others' belief states through unequal distribution of information among individuals, one expects to elicit cognitive representations conducive to experimental investigations into the attribution of beliefs to others (i.e. ToM). If task conditions favoring the formation of cognitive representations of others' beliefs are weak or absent, then detection or discrimination between participants' own belief states and participants' representations of others' belief states becomes impossible. In addition to divergent belief states, the perspective taking task by Baker et al. (2017) includes dynamic belief updates. As the observed agent moves around the environment, more information about the possible goal states becomes available and the observed agent's belief is updated. To correctly predict behavior, participants have to track these belief updates leading to an alignment of their own belief and their representation of the observed agent's belief. That means, participants not only have to represent the agent's belief but also to dynamically update these representations. Additionally, participants encounter social uncertainty in this task. The observed agent's preferences are unknown and need to be inferred from observed behavior. This adds a second inferential process. In addition to updating the other agent's beliefs about the environmental context based on the observation of the agent's behavior, participants must also infer the preferences of the other agent. Thereby, a manifold intentional representation of the observed agent is generated, creating a scenario that allows examining the attribution of beliefs and preferences at the same time.

However, while in both tasks the observed agent's intentional states are highly relevant, participants themselves take a purely observational perspective. They are detached and removed from the scenario and their judgments and predictions are entirely inconsequential to the characters and the progressions of the scenes they observe. Thereby, the prediction process taking place in the observed agents' reference frames is disconnected from participants' self-referential cognitive processes, with the possible exception of a participant's motivation to give accurate answers and not to be seen to be in error.

2.1.2. False belief reasoning and perspective taking in individuals with AS

Tasks that use observation under asymmetric distribution of information between observer and observed agent inform most studies of children and adults with AS. In these tasks, children with AS often fail to show accurate explicit false belief reasoning (Baron-Cohen et al., 1985). This is sometimes interpreted as children with AS failing to have ToM. However, as Premack & Woodruff (1978) noted, having ToM means positing others' motivations, intentions, goals and beliefs and does not necessarily entail having accurate ToM. As Gernsbacher & Yergeau (2019) show, there is no strong empirical evidence in favor of claims that children, adolescents, and adults with AS lack ToM. Studies making such claims generally point to the impaired accuracy rather than complete absence of ToM. Several findings indicating a specific impairment in ToM in AS might actually be potentially related to more general impairments, as IQ is a strong predictor of performance on ToM tasks in AS (Buitelaar et al., 1999). While most recent studies ensure that all participants with AS have an IQ that is at or above "average intelligence" (i.e., IQ > 85; e.g., Vivanti & Rogers, 2011), some studies that claim ToM differences continue to demonstrate differences in IQ that are not controlled for in analyzing group differences (Assaf et al., 2013). Senju et al., (2010) argue that standard explicit false belief tasks require such high verbal and other cognitive abilities that they may yield false positives on ToM impairment, at least in the case of spontaneous ToM. To control for verbal intelligence differences, Senju and coworkers (Senju et al., 2009) used a passive viewing false belief task. They demonstrated differences in anticipatory eye movement between NT adults and persons with AS, which they interpreted as indicating impaired implicit ToM in adults with AS who demonstrated accurate ToM via explicit false belief tasks.

Perspective taking tasks similar to Baker et al. (2017) were used to test visual perspective taking abilities. Level 1 visual perspective taking (VPT1) is the ability to accurately tell whether another agent is able to see an object or a feature of an object or not. Level 2 visual perspective taking (VPT2) denotes the ability to understand that two agents might see the same object differently. Thus, VPT2 focuses on how the same object is perceived by different agents (Pearson et al., 2013). Pearson and colleagues (2013) reviewed several papers examining VPT1 and VPT2 in AS. Most studies of VPT1 reported no differences in AS compared to controls. VPT2 differences in AS were inconclusive as studies reported conflicting results. In a study of implicit VPT1, Cañigueral & Hamilton (2019) found that adults with AS showed no preference in looking at recorded video clips of agents who could see vs. not see. Controls preferred video clips in which the agent could see, suggesting that social gaze in controls but not in adults with AS may not be influenced by implicit ToM.

2.1.3. Perspective taking in nonhuman primates

Since 1987, when Premack and Woodruff asked whether a chimpanzee has a Theory of Mind, experimental investigations have generated competing evidence for and against ToM in nonhuman primates. In a review summarizing research from the 30 years following this initial question, Call & Tomasello (2008) conclude that chimpanzees do understand others in a perceptual-goal perspective taking task but fail to represent them as full intentional agents with beliefs and desires. However, experimental evidence is variable and inconclusive. Experimental studies of nonhuman primates often use observational and perspective-taking tasks under asymmetric environmental uncertainty, starting with the work reported by Premack and Woodruff (1978). Tomasello and coworkers (Krupenye et al., 2016) used recorded video clips and measured anticipatory looks by bonobos, chimpanzees, and orangutans to assess ToM for goal-directed behavior by observed human agents. Most apes showed gaze that anticipated that human agents would act by those agents' false beliefs. In his 2007 review, Premack broadened the false belief tasks to include direct gain and loss relevance to participating primates, emphasizing the importance of social engagement via direct consequential outcomes for the observer/participant. For example, nonhuman primates will wait until a human is not looking before attempting to obtain food that is within reach but not yet offered. In an interactive task testing level 1, visual perspective taking, home-reared chimpanzees showed strong evidence for understanding which of two human assistants had sight of a valued food object. Similarly, Tomasello and colleagues (Karg et al., 2016; Schmelz et al., 2011) used turn-taking, semi-interactive tasks to conclude that while it is unclear whether chimpanzees are capable of second level visual perspective taking, they do understand that conspecifics can use visual information to infer consequential state variables such as the location of a valued food item. Frans de Waal and colleagues (Hall et al., 2017) also used semi-interactive tasks with consequential outcomes in collecting evidence that chimpanzees can engage in tactical deception and can recognize that conspecifics do so as well. Despite the richness of these findings, their interpretability has been questioned. Some scholars reason that meaningful examination of ToM in non-verbal species would require clear definitions of what non-verbal representations of other look like, clearer operational definitions regarding convincing evidence for ToM, and how this all could be realized experimentally (Penn and Povinelli, 2007).

2.1.4. Observational learning tasks

Albeit not interactive, observational reinforcement learning paradigms require participants' active engagement because their own gains and losses, and potentially those of others, are at stake. Experimental setups examining observational learning deploy classic single-agent reinforcement learning problems such as probabilistic reversal learning or multi-armed bandit tasks. In these tasks, the goal is to choose optimally among multiple competing choice options. However, participants receive only probabilistic information about choice-outcome associations, so choosing optimally requires dynamic learning about the environment. During observational learning, instead of choosing actions and receiving outcomes themselves, participants observe other individuals selecting between available options for rewards (Hill, Boorman, & Fried, 2016; Selbing & Olsson, 2017). From observed choice-outcome associations, participants vicariously learn the underlying reward distributions. In such settings, the observed agent's choices determine the information the observer receives about the environment. Apart from that, the observed agent's actions and consequently his representations of the learning problem are irrelevant to the observer. However, work on preferences alignment has shown that participants that observe others' preferences in a decision task are influenced by those traits, even if these are not directly relevant for the decision problem at hand. Using a social variant of delay discounting, a task that requires arbitration between smaller immediate and larger future rewards, Moutoussis et al. (2016) showed that participants' preferences aligned with others' after observing their choice behavior. Devaine and Daunizeau (2017) interpret such attitude alignment as adaptive behavior in difficult decision problems as it provides additional information on how to react to a highly uncertain decision situation.

Extending vicarious learning into a more immersive setting, Zhang & Gläscher (2019) examined the effect of observing multiple other learners while engaging in a probabilistic reversal learning task oneself: participants choose between two choice alternatives, one associated with a high probability of winning, the other with a high probability of losing. Reward contingencies switched after a variable number of trials. After making their own first choice, the other learners' choices were revealed and participants received the opportunity to adapt their decision, i.e., switch to the other option or stay with their initial decision. Thereby, the task allows for learning from one's own and others' experience (Figure 1C). Zhang & Gläscher found that the stronger the social information diverged from participants' own choice (i.e., the more co-players chose the opposite option to themselves), the more frequently participants switched their choice to go with the group.

A second engaging observational learning paradigm was introduced by Boorman and colleagues (Boorman et al., 2013): the task mimics a stock market scenario, in which observed agents had to predict different assets' changes in value. Participants observed the agents' learning success and were asked to bet for or against their choices while the observed agents completed the task with varying success. Figure 1D shows an exemplary situation with four agents (green, purple, blue and orange) with different underlying fixed success rates

(0.3, 0.4, 0.6 or 0.7) that are unknown to participants. In addition, participants themselves had to occasionally predict the assets' value development. To perform well in this task, participants needed to track the properties of a volatile environment (i.e., the value of assets) as well as the others' expertise in the task, i.e., the overall correctness of their choices. Computational modeling results indicated that participants tracked the observed agents' task abilities in a two-step process: first, during the choice period they evaluated the observed choice in light of their own estimate of the asset's value (i.e., their own representation of the environment). Second, at outcome presentation they updated their estimate based on the observed agents' success rates.

The last variation of observational RL discussed here motivated tracking agents' learning by hiding action outcomes while the observed agents' preferences were known to the participants (Collette et al., 2017) (Figure 1E). Participants inversely inferred their own subjective action values from the choices and preferences of the observed agent, assuming that they would act to maximize their reward. They learned about the environment by interpreting observed behavior via a representation of the respective agent's value space and integrating own preferences with this information to perform actions.

Observational RL tasks instruct participants to maximize rewards in highly uncertain environments where information is gathered by observing own and/or others' actions and outcomes. The need for intentional representations of others varies across those tasks: in vicarious RL observed actions are merely used to track environmental properties without the need for an explicit intentional model of another agent. In social influence tasks, other's actions are relevant for one's own decision, but the other agent's frame of reference is irrelevant, whereas in inverse RL tasks, the other agent's action have to be inferred based on his known preferences.

2.1.5. ToM processes in observational tasks

False belief (Wimmer and Perner, 1983) and perspective taking (Baker et al., 2017) tasks require no choices from participants but require judgments that depend on the representation of observed agents' perspectives on the environment. Most importantly, information about the context is distributed unequally between the observer and the observed agent, eliciting divergent knowledge states. This divergence makes the other's knowledge state relevant for the observer's predictions and judgements. However, as no choice or action that would contribute to the dynamics of the scene is required, this reasoning remains detached from the observing participants' own frames of reference. False belief and perspective taking tasks therefore require representations of others' intentional states but no integration of these representations with self-referential cognitive processes.

In observational learning tasks, participants make choices after observing other agents' behavior in an uncertain environment. In vicarious RL (Hill et al., 2016), observed actions determine the observations about the probabilistic reward structures. Similarly, in the influence task (Zhang and Gläscher, 2019), the group's choices can be interpreted by participants as information regarding the quality of their own choices. In the expertise tracking task (Boorman et al., 2013), uncertainty about the observed agent's competence is added. The observed actions have to be evaluated in relation to participants' own world knowledge and the estimated expertise. However, modelling results indicate that participants did not assess expertise from the other agents' perspective but evaluated choices in their own valuational frame. In these tasks, it is irrelevant how observed actions came about. Hence, it is likely that participants integrated observed behavior and action-outcomes into their own representations of the environmental states and transitions, but others' perspectives and the intentional states leading up to their actions were irrelevant to them. This is different in the inverse learning task by Collette et al. (2017). This experimental setting requires participants' emersion into the other agents' learning processes, potentially eliciting an intentional representation of others' intentional states and evaluation of these representations with respect to the participants' own preferences.

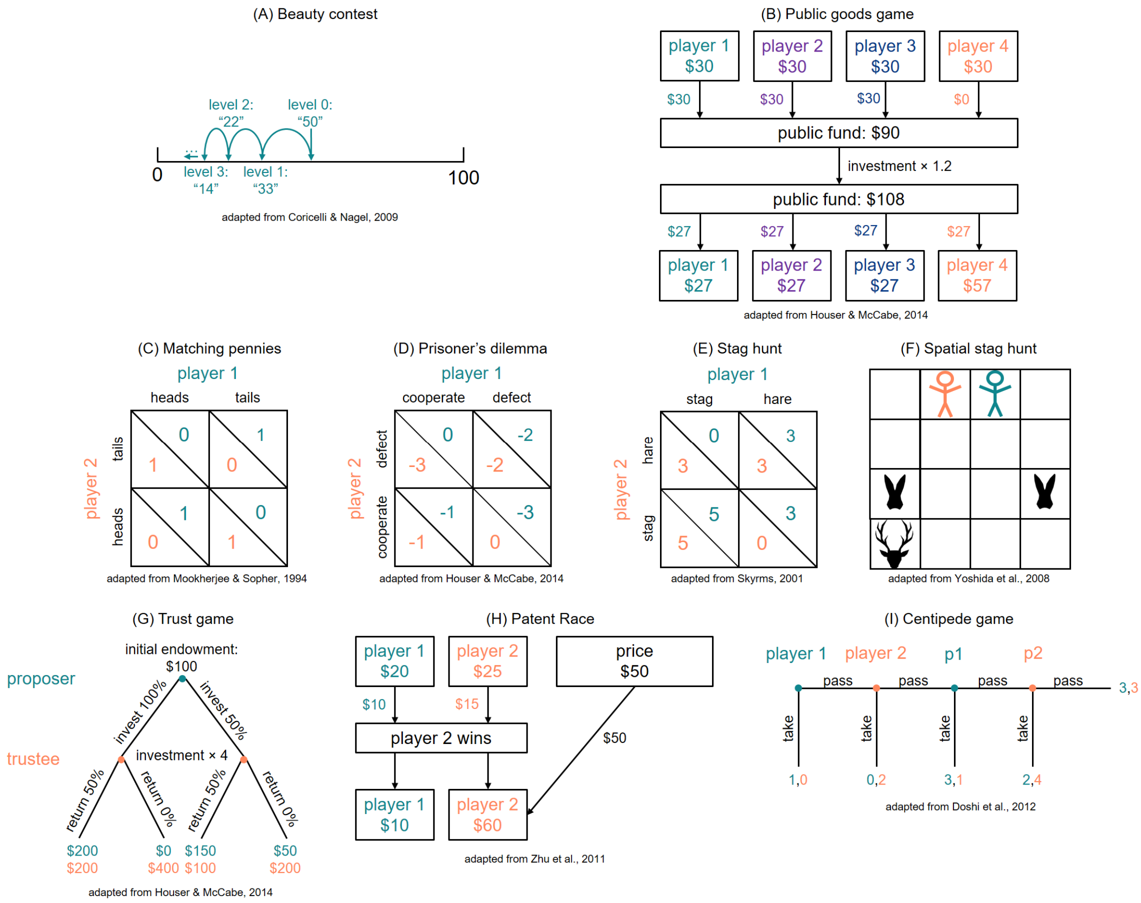

Figure 2| Interactive tasks in stable and fully observable environments. **(A)** The beauty contest game nicely illustrates the concept of recursive reasoning. The goal is to choose a number between 0 and 100 that is closest to 2/3 of the average of the numbers chosen by all other participants. It is assumed that depending on their level of reasoning, players choose different values: level 1≈33, level 2≈22, and so forth. **(B)** In the public goods game, a group of players is endowed with an initial amount. They can choose to invest as much as wanted into a public fund. The fund is then multiplied by a fixed factor and equal splits are returned to all players irrespective of their initial investment amount. Additionally, players get to keep the money they did not invest into the public fund. **(C) (D) (E)** Matrix Games, such as matching pennies, prisoner's dilemma and stag hunt are defined by a payout matrix. The payout matrix determines both players' rewards based on the two players' actions. Depending on the configuration of the matrix a competitive or cooperative coordination scenario is created. **(F)** Grid games like the spatial stag hunt add a spatial component to games defined by simple payout structures as in the previous examples. To successfully coordinate in the spatial stag hunt, players need to take the path their co-player is taking through this environment into account; hence inference about the co-player's future actions is added to the decision process. **(G)** In the trust game one player acts as the proposer, the second as the trustee. Actions are taken sequentially. The proposer can decide how much of an initial endowment to invest. The investment is multiplied by a known factor. The receiver can now decide how much of the multiplied investment to return to the proposer. **(H)** The patent race game comprises two players, a rich and a poor player. Both players can choose how much of their capital to bid for a price. The higher bid earns the price, both players' bids are lost. **(I)** In the centipede game two players sequentially choose between "take" or "pass". When a player takes, unequal rewards are distributed to both players, and the taking player receives more. Importantly, with each move, rewards increase. However, the player whose turn it is gains a greater reward than that the waiting player.

2.2. Interaction under full environmental certainty

When social scenarios expand beyond mere observation to direct interaction, interacting individuals' behaviors and consequently successful outcomes depend more strongly on their thought processes becoming interdependent. This means, in situations in which an agent A's actions are relevant to a second agent B while also B's actions are relevant to A, their reasoning processes may become recursive (in the sense of "A thinks that B thinks, that A thinks that B will do XYZ.") (Camerer, Ho, & Chong, 2004). This increase in interdependent, or higher level, ToM is even more likely the more that successful task outcomes depend on higher level ToM. To examine reasoning processes of this kind, behavioral game theory uses a range of simple yet very powerful interactive tasks, generally called "games". A game in this sense is a multi-agent decision situation where the actions of all participating agents affect each other. That is, each individual's success (usually defined as maximizing the individual reward) depends on others' choices. Assuming complete and optimal rationality of all interacting agents (Gibbons, 1992), the field was initially concerned with computing optimal solutions to such games. These solutions are generally termed equilibrium states. Deviating from these equilibrium states would be detrimental for all agents. The most famous example of such a state is the Nash equilibrium. However, more recently, experimental economics and behavioral game theory have focused on more descriptive rather than exclusively normative questions, exploring which factors affect actual human social decisions, which are often suboptimal from the perspective of rational choice theory (Camerer, 2003).

Recent reviews summarize a multitude of studies that deploy a variety of strategic games and examine the effects of instructions, framing, incentives, and many other highly relevant factors driving human interactive decision-making (Houser and McCabe, 2014). Here, we can neither list all existing games, nor summarize the effect of the different factors affecting strategic behavior listed above. We merely pick a subset of characteristic economic games and discuss them with respect to our two defining dimensions: interactivity and uncertainty. Furthermore, although classic economic research often tests strategic behavior in one-shot scenarios, we only consider recurring interactions. In everyday life, humans tend to interact with the same individuals more than once. Moreover, repeated interactions allow for sophisticated predictions about others' behavior based on regularities in their strategies or on built-up expectations about their motivational states.

2.2.1. Anonymous group interactions

The beauty contest game (Figure 2A) illustrates the concept of recursive reasoning in more detail: a group of individuals anonymously chooses between a number between 0 and 100. The goal in the task is to choose the number closest to 2/3 of the average of the numbers chosen by all players. According to cognitive hierarchy theory (Camerer, Ho, & Chong, 2003), people engage in reasoning processes of varying levels of sophistication to solve this problem: A very basic player, so called level 0, randomly chooses a number. A more sophisticated level 1 player assumes all others to act at level 0, resulting in an average of 50, and chooses 33 (2/3 of 50). A level 2 player considers the others as level 1 players yielding and optimal response of 22 (2/3 of 33), and so on. Such recursive reasoning could in principle extend ad infinitum, leading to the optimal equilibrium of 0. Nevertheless, on average, people select numbers between 25 and 40 suggesting a reasoning level of 1 or 2 in this task setting, but the variance in choices is large and groups of varying analytical training (e.g. those with a PhD in economics compared to high school students) show highly different means (Camerer, 2003; Camerer, Ho, & Chong, 2015).

The public goods game (Figure 2B) deploys anonymous group decisions to study fairness and reciprocity in a situation where narrow individual interests may conflict with the gains or losses across an entire group (Houser and McCabe, 2014; Khalvati et al., 2019). All members of a group are endowed with the same initial amount on each trial. They may choose how much of their individual money to secretly invest in a public fund. The money in this fund is then multiplied by a known factor and equal shares are distributed to all players irrespective of whether they gave or chose not to give to the public fund. Players may choose to keep all of their money on a trial, in which case they have their initial endowment plus their equal portion of the group distribution from the public fund. This means players can choose to cooperate by investing into the public fund, but they can also “free-ride” by taking their share of the others’ investment without investing themselves, maximizing only their individual reward. However, free-riding is known to reduce the group’s overall investment in the public fund, especially as more people in the group choose to free-ride as they see others free-ride. This means that everyone, including the free-riders, receives a lesser distribution. Consequently, players need to consider their own actions’ effect on the group’s behavior to maximize their outcomes. Overall, people invest about half of the initial endowment although the cooperation rate drops over repeated interactions if the group members remain the same (Ledyard, 1994). Free-riding for a given trial also increases with increasing numbers of free-riders on the previous trial. But not all participants show this pattern. Participants who conditionally cooperate or go beyond reciprocation generally show greater attention to others’ preferences and expectations, especially over multiple trials involving the same group members (Chaudhuri, 2011). Yamakawa and coworkers (Yamakawa et al., 2016) partnered a participant with a computer on the public goods task. Participants were informed that the amounts they gave would have no effect on the computer’s predetermined investment into the public fund, and so no effect on the participant’s own gains. Under these conditions of full lack of interactivity and full environmental certainty, participants exhibited near 100% free-riding behavior. Fischbacher & Gächter (2010) showed that conditionally cooperating participants in the public goods task were sensitive to the heterogeneity of agent preferences. Computational models predict that greater environmental uncertainty in public goods tasks could elicit greater sustained cooperation through attention to the preferences of others in the group (Kurokawa and Ihara, 2009).

Beauty contest and public good games immerse participants in an interactive scenario, where actions of all involved individuals directly affect each other. However, as the interaction takes place at the anonymous group level, effectively assessing how participants represent others as distinct individuals with individually different reasoning processes is not possible. Instead, time-varying average group level variables are the target of analyses. For example, in the beauty contest, one estimates the group’s level of sophistication. In public goods games, one estimates the group’s preferences relating to investment. This requires observing and learning about these variables as interactions continue and so resolving social uncertainty. Hence, instead of explicit recursive reasoning about other individuals’ specific cognitive processes, participant models focus on the recursive reasoning about the group as a whole.

2.2.2. Dyadic games

Dyadic games (i.e., those involving two players) allow for more direct interaction and individually focused ToM than do anonymous group tasks. First, we consider strategic two-player games entirely defined by a single payoff matrix: conditioned on both players’ actions a payoff matrix alone determines individuals’ rewards. Well-known examples comprise “matching pennies” (Mookherjee and Sopher, 1994; Touhey, 1974) (as variant also known as “hide and seek”, “inspection game”, or “rock, paper, scissors”), “prisoner’s dilemma” (Axelrod and Hamilton, 1981) (Figure 2C to E), and “stag hunt” (Rousseau and Cranson, 1984; Skyrms, 2001). Matching pennies is a zero-sum game fully defined by a competitive payoff structure. One player's win determines the opponent's loss and vice versa (Figure 2C). Stag hunt constitutes a cooperative game which requires coordination between partners to obtain the highest possible rewards: jointly going for the large reward (i.e., both hunting the stag instead of going for smaller individual reward, the hare) yields the maximal payoff for both players (Figure 2E). The famous prisoner's dilemma incorporates both, competition and coordination. Both players have the option to "cooperate" or "defect". When joint actions are uncoordinated (i.e., one player choose cooperates and the other defects), the defecting player gets the best outcome, while the cooperating player receives the worst outcome (Figure 2D). Coordination (both players choose identical actions) averts one's big loss and leads to a symmetric, but suboptimal outcome for both players.

Spatial variants of these simple matrix games, also referred to as grid games, have been created by adding two-dimensional grid worlds to the elementary payoff structure of fully cooperative and social dilemma type of games (Kleiman-Weiner et al., 2016; Shum et al., 2019; Yoshida et al., 2008). In these grid worlds, the underlying payoff structure of a game is preserved, e.g. in a spatial stag hunt jointly catching the stag yields the highest reward (Figure 2E), but additionally coordinated long-term action planning is required (Figure 2F). The rate of cooperation in grid games depends on the underlying payoff structure, such that there are higher cooperation rates in pure coordination games and lower cooperation rates in dilemma type of scenarios (Kleiman-Weiner et al., 2016). Yoshida and colleagues showed that cooperation rates also depend on the strategy of each partner: They found higher cooperation rates with partners of higher sophistication level (Yoshida, Seymour, Friston, & Dolan, 2010; Yoshida et al., 2008).

2.2.3. Bargaining games

The last group of interactive games considered here, including the "trust", "patent race" and "centipede" games, consists of simple bargaining environments (Sanfey, 2007) (Figure 2G to I). In the trust game (Berg et al., 1995), one of two players, the investor, is endowed with a certain amount of money (Figure 2G). The investor decides how much of that amount is to be invested with the other player, the trustee. Upon investment, the money is multiplied by a fixed factor and the trustee can then decide how much of the resulting amount to return to the investor. When the investor invests a lot and the trustee returns a fair share of the multiplied investment, from the perspective of both players, both players mutually benefit. However, if the trustee does not return at least the invested amount, the investor loses money and a cycle of distrust begins. This tends towards the investor making investments of less money, which lowers payoffs for both players. In a multi-round trust game, players commonly follow a tit-for-tat strategy. They cooperate when the co-player cooperated in the previous round and likewise do not cooperate (lower the amount of monetary transfer) following non-cooperative behavior. But both, investments and returns, slightly reduce over time (King-Casas et al., 2008, 2005). However, some trustees show "coaxing" behavior when investors' investments decrease substantially and they return a larger share of the multiplied investment to reassure the investor of their trustworthiness (King-Casas et al., 2008). These findings indicate that behavior in the trust game relies on reasoning processes on how actions affect a co-player's impression of one's own trustworthiness and cooperativeness. In the patent race game (Dasgupta and Stiglitz, 1980; Loury, 1979), two players, competing for a prize, receive initial endowments (Figure 2H). However, one player receives more and is "richer" than the opponent. Both players simultaneously bid for the price. The player that offers the larger amount wins the prize but loses his investment, while the second player loses his investment and also the prize. To maximize their returns, players need to invest as little as possible but as much as necessary to outbid the opponent. Based on their opponent's choice history, participants can make predictions of their opponent's next offering. Further, players can assume that their partner is predicting themselves, triggering recursive reasoning processes.

In the centipede game (Rosenthal, 1981) (Figure 2I), two players take turns at choosing between keeping a pot of money or passing it on to the co-player. If the first player passes the money on and the co-player keeps in the next round, the first player's outcome is slightly lower than if he would have kept it. However, after a round of passing by both players, both players' outcomes increase. As in the trust task, the centipede game requires reasoning about how one's own behavior will affect the co-player's future actions, especially whether that person will reward or punish mistrusting choices, respectively. Additionally, participants could engage in representing and reasoning about their partners' representation and reasoning about them, and so on. For instance, a fully rational and optimal player solves the centipede game by backward induction: "in order to receive the maximum reward at the end, the other person has to see the advantage of passing, which implies that I have to pass also," etc. Such reasoning quickly reveals that the optimal solution is to take the money in the first step, thus insuring at least a minimal reward. However, humans often advance the game to a later stage (Hedden and Zhang, 2002; McKelvey and Palfrey, 1992) potentially due to their limited capacity for complete backward reasoning all the way through (Ho and Su, 2013) or because they recognize the mutual long term benefit of reciprocating and/or altruistic behavior.

2.2.4. ToM processes in interactive games

This brief overview of the characteristic features of game-theoretic tasks indicates how interdependent decision processes are favored in interdependent designs. The ability to strategically respond cooperatively or competitively in these tasks requires forming a representation of another agent's motivational states and reasoning processes and types, e.g. whether the other person follows win-stay-lose-shift, tit-for-tat, choses actions at random, etc. (Axelrod and Hamilton, 1981). Individual state and trait variables such as risk aversion, ambiguity aversion, fairness, trust, and greed come into play in dilemmas and bargaining tasks (Engel and Zhurakhovska, 2016).

We can perceive the dimensions of uncertainty and interactivity in these tasks, and map those onto how likely they elicit ToM at all as well as how likely they elicit deeper levels of recursion in ToM. In all interdependent tasks, irrespective of the specific payoff structure, whether they are cooperative or competitive, players' choices and outcomes are bound together in real time, thus favoring active recursive reasoning on a trial-by-trial bases. Further, if one's own and the others' interests diverge, evaluating joint decisions requires greater cognitive effort (Emonds et al., 2012). All tasks discussed in the previous section are strong in attempting to maximize social uncertainty. In the fully cooperative stag-hunt game, for example, both players' overall goals can be assumed to be aligned given the cooperative setup. Therefore, motivational variables tend to be less important in this scenario, but uncertainty emerges from individuals' risk aversion (Büyükboyacı, 2014). However, these tasks do not present participants with uncertain environmental contexts or include uneven distribution of information between agents. Consequently, while success in these tasks depends on representing others' motivational states and individual character traits, etc., none of these tasks requires updating representations of others' beliefs about the states or situations constituting the environment. To drive participants to engage in ToM more consistently over the course of a task, and to represent of others' motivational and knowledge states, we need tasks that expand on both social and environmental uncertainties.

2.2.5. Strategic interaction in persons with Autism Spectrum

The vast majority of investigations into ToM abilities in autism spectrum (AS) has been conducted using false belief and perspective taking tasks. However, more recently strategic games and computational cognitive modeling have been deployed to examine strategic reasoning abilities and differences in individuals with AS. In a "reverse Turing-test" (D'Arc, Devaine, & Daunizeau, 2018) persons with AS and neurotypical (NT) controls played "hide and seek" (similar to matching pennies, Figure 2C). In this competitive game one individual wins, if the opponent's choice mismatches. For example, if a participant hides behind a tree and the opponent searches behind the wall, the participant wins. Using social and non-social framing, D'Arc and colleagues (D'Arc et al., 2018) showed that strategies of persons with AS did not differ between a supposedly human or computer opponent, while NTs successfully competed in the social but not in the non-social condition. Further, persons with AS successfully competed against fictitious opponents modeled using random and simple fictitious play (a strategy based on the opponent's preceding choice frequency) but failed in competition against opponents following higher level recursive models. These results support findings from the first quantitative examination of recursive social reasoning in AS by Yoshida et al. (2010). Using the spatial stag hunt game (Figure 2F) with adults with autism, they found that some individuals with AS exhibited extreme choice behavior, such that they never cooperated or never competed. This patter was absent in NTs. The authors showed that the severity of AS symptoms correlated with AS participants' abilities to successfully compete against recursive decision strategies. These first studies quantifying strategic behavior in persons with AS using computational models provide fine-grained insight into the ToM differences during strategic interaction in AS.

2.2.6. Strategic interactions in nonhuman primates

A focus on ToM tasks developed especially for and applied to understanding ToM in nonhuman primates is necessary for a full understanding of how ToM relates to evolution in simiiformes, especially to the roles of cooperation and competition in the evolution of hominids and hominins. Coordinated cooperation and competition appear to have played critical roles in the evolution of simiiform brains and intelligences. Within the social brain account (Dunbar, 2009), which is the hypothesis that primates' large brain volume and complex social cognition developed in response to the demands in increasingly complex social groups, cooperation and affiliative group bonding are more prominent as the driving forces in cognitive and brain evolution. This view has support from accounts that argue for a more important role of pro-social behaviors in facilitating in group bonding (Barrett and Henzi, 2005). On the other hand, the *Machiavellian* hypothesis (Whiten and Byrne, 1988) focusses on the evolution of sophisticated ToM by emphasizing competitive interactions and the need to outperform others in the competition for resources.

Little is understood about the cognitive and neural systems underlying coordinated cooperation and competition. In the absence of computational evidence, there are ongoing debates about whether evolutionary costly expansions of the primate brain owe more to increased need for cognitive resources, including ToM, to cooperate (Dunbar, 2009) or to manipulate and dominate (Byrne, 2018). Recent computational approaches, drawing on several of the task designs reviewed here, lend insight to these questions about the role of ToM in evolutionary history. Devaine and colleagues (2017) examined seven different nonhuman primate species' responses in hide and seek games. It was found that all species showed less sophisticated behavior than humans and that ToM abilities varied with species' overall brain volume but not social group size. The authors suggest these findings support a general intelligence rather than a social brain hypothesis, such that the evolution of sophisticated ToM was determined by overall increasing cognitive abilities rather than in response to the increasing complexity of social communities. Comparing the behavior of chimpanzees and humans in a competitive inspection game, Martin and colleagues (Martin et al., 2014) found chimpanzees' choices to be much closer to equilibrium (i.e. optimal behavior according to norms of rational choice theory), than the choices made by humans. Chimpanzees followed rational choice theory while humans depart from it in favor of more cooperative choices. This could be due either to a greater propensity for cooperation in humans when interacting in small groups with relatively low stakes, to the fact that humans depend on language to make optimal choices, or to some combination of both of these possibilities.

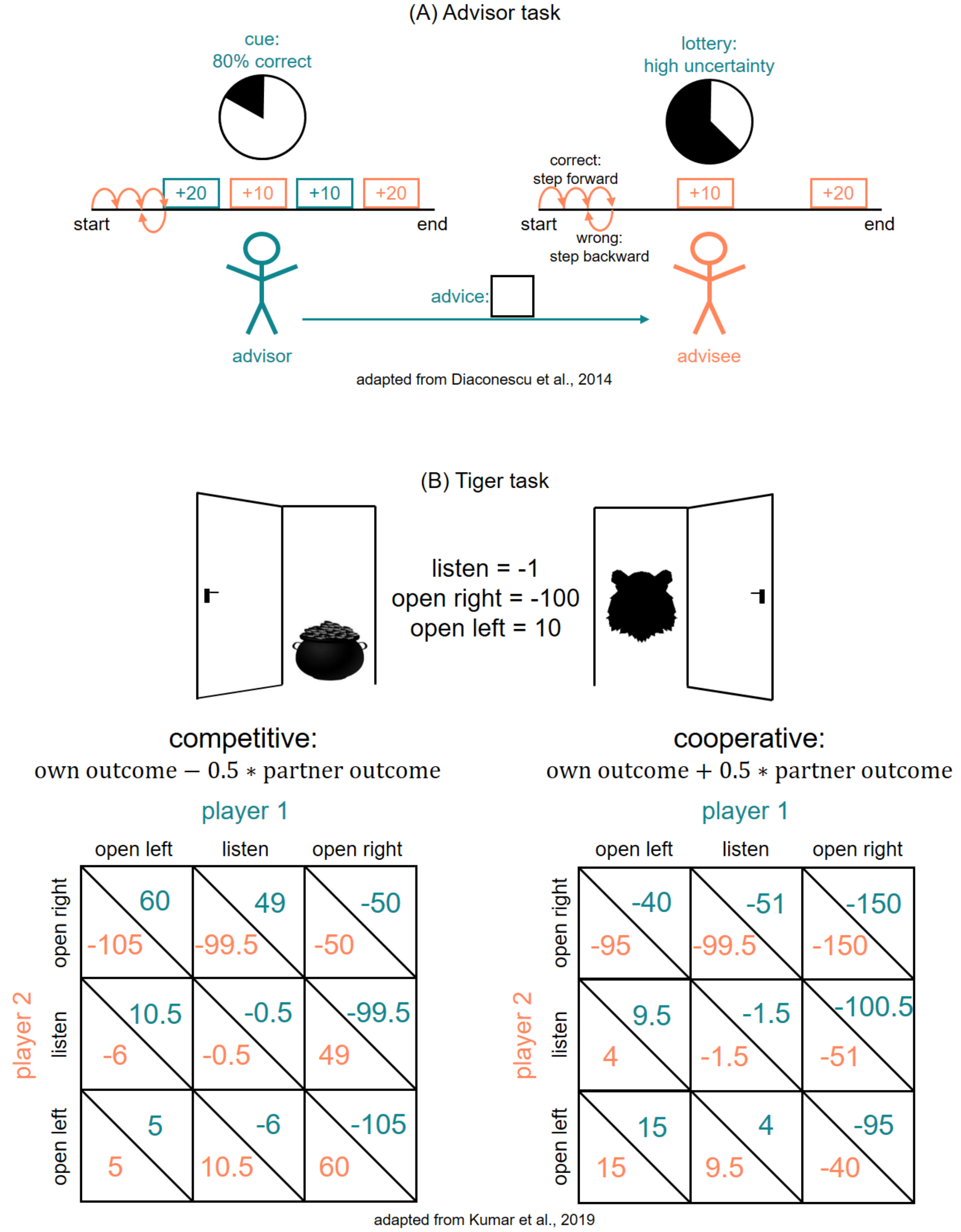

Figure 3| Interactive decision tasks under uncertainty

(A) In the advisor task one participant takes the role of the advisor, the second of the advisee. The advisee has to make choices on a probabilistic lottery with high uncertainty. Correct choices move the advisee forward on a progress bar, incorrect choices send the advisee backwards. Additionally, if the advisee finishes in two predefined regions, the advisee receives a bonus of +10 or +20 but not the advisor. The advisor receives more accurate information about the outcome of the lottery and can choose to send advice to the advisee. However, the advisor's bonus regions are defined differently than the advisee's bonus regions, creating a conflict via competing interests. **(B)** The tiger task simulates a simple game show scenario: two players are faced with two doors. Behind one is a pot of gold (a positive reward of +10) behind the other door is a tiger (a large negative punishment of -100). The players have to choose between three choice options: (1) listen (providing probabilistic information about the tiger's location at a small cost of -1); (2) open the left door; or (3) open the right door. Depending on the reward configuration, competitive (left) or cooperative (right), co-players have to race to identify and open the door of the gold or they have to coordinate responses in identifying and opening the door to the gold, respectively.

2.3. Interaction under social and asymmetric environmental uncertainty

Very few experimental approaches to date have combined asymmetric distribution of information, environmental uncertainty and interactivity. One example is the advisor task (Behrens et al., 2008; Diaconescu et al., 2014) (Figure 3A) in which two participants, an advisor and an advisee, interact in a gambling task. The advisee has to choose between two uncertain lotteries, while the advisor, who is provided with more accurate information on outcome probabilities, can send simple cues to the advisee on which action to take. After a correct choice the advisee moves forward on a progress bar and receives a small reward, but an incorrect choice sends the advisee backwards and results in a small financial penalty. When the advisee moves into pre-specified regions on the progress bar associated with either a bonus for the advisee or the advisor, there is an additional payout of $10 or $20 for one or the other, but not both. These regions are fully disclosed to the advisor, but the advisee only knows their own bonus regions, while being ignorant about those of the advisor. These different regions are designed to manipulate the motivational states of the advisor. During one phase of the game, the advisor acts to ensure that the advisee moves into one of the advisee's bonus regions, providing veridical advice about the outcome probabilities. Other phases of the game incentivize the advisor to hinder the advisee from reaching the advisee's own bonus region by providing false advice. A participant who uses no ToM in the task would ignore the advice and would simply need to model the likelihood that the lottery corresponds to the true outcome (i.e. environmental uncertainty). A participant that takes into account the asymmetric distribution of information about the lottery between advisee and advisor, and advisor's and advisee's competing interests is likely to use ToM. Modeling results by Diaconescu and collegues (2014) indicate that participants use ToM in the advisor task. They found that choices made by participants that took the role of the advisee were best explained by a model that included a parameter estimating the advisor's current tendency to be accurate and a parameter estimating the advisor's likelihood of deception across the trials (i.e., social uncertainty). Participants did better when they estimated advisors to be reliable and accurate, and in fact advisors gave accurate information about 75% of the time.

A powerful experimental setting combining full interactivity and uncertainty is the multi-agent tiger task (Doshi and Gmytrasiewicz, 2004; Kumar et al., 2019) (Figure 3B). In a scenario that mimics a game show setting two players have to learn which of two doors hides a pot of gold (reward of +10) and which hides a dangerous tiger (punishment of -100). They can choose to open one of the doors or sample probabilistic information about the location of the tiger and about the other player's action at a small cost (-1). If a door is opened, the tiger's location is randomly reset, and the game starts anew. A superimposed reward structure incentivizes players to cooperate or compete. In the cooperative setting, after each action players receive half of the partner's outcome while in the competitive scenario half of their partner's outcome is subtracted from their own outcome. In other words, under competitive conditions, an opponent's loss results in a win for oneself, under cooperative conditions the partner's loss results is also one own's loss. Additionally, periods of divergent knowledge states between co-players occur when one chooses to sample more information while the second opens the door. An opening action provides a short window of more information. First, the co-player's choice is revealed. Second, after opening the door, the player finds out where the gold is and knows that the location of the gold and the tiger will be reset, so that previously sampled information needs to be discarded. This asymmetric knowledge about the state of the gold and tiger is an advantage in the competitive setting but detrimental for the cooperative condition. In the competitive setting players race to open the correct door before the opponent does, while they still need to sample enough information to avoid the tiger. The cooperative scenario incentivizes coordinated behavior over individual learning about the reward distribution. This typically results in fewer trials spend sampling the information in the competitive setting, which leads to riskier choices, but also has the potential to beat the opponent to the gold (Kumar et al., 2019).

2.3.1. ToM processes during interaction under social and environmental uncertainty

Both the advisor and multi-agent tiger tasks include asymmetric distribution of information and high environmental and social uncertainty. All participants receive only probabilistic information about the state of the environment. This means not only they themselves form uncertain beliefs but also can only estimate the other's belief about the state of affairs. Additionally, in the advisor task, the advisor receives more veridical information than the advisee, creating divergent belief states between advisor and advisee. In the tiger task beliefs diverge when players perform different actions and one player receives more information than the other. As in the false belief task, divergent beliefs in advisor and tiger task create an incentive to represent the other's knowledge state. However, the false belief task is purely observational. In contrast, in the advisor task advisees react, and in the tiger task players fully interact. Further, advisor's cues in the advisor task add additional uncertainty and the advisee needs to infer the advisor's intent and trustworthiness. Advisees need to integrate both their prediction about the advisor's trustworthiness and truthfulness in giving advice on a given trial, which in turn results from their representation of the advisor's intentions, together with their own beliefs about the lottery. In contradistinction, incentives are clearly set in the tiger game. However, individuals may differ with respect to riskiness. Players need to estimate and react to their co-players' individual levels of risk-seeking behavior and integrate that into the rest of the task components. These task properties likely trigger complex social reasoning processes including a representation of the other person's dynamically changing beliefs and their motivations. Players have to integrate their estimate of the co-player's learning, valuation and decision process with their own. Additionally, they need to consider how they themselves are represented by the other. Hence, one's own reference frame and that of the other are recursively interweaved. By combining environmental uncertainty and unequal distribution of information about the environment these tasks promote and therefore allow careful computational analysis of the attribution of dynamic belief states. Uncertainty about others' intent and risk aversion requires additional inference about motivational states. Finally, incorporating interactive settings creates opportunities to examine the dynamic attribution of trial-by-trial changes in others' intentional states, the prediction of behavior in light of these states, active choice with respect to one's own knowledge and motivations, and reflections about one's own reference frame from others' perspectives.

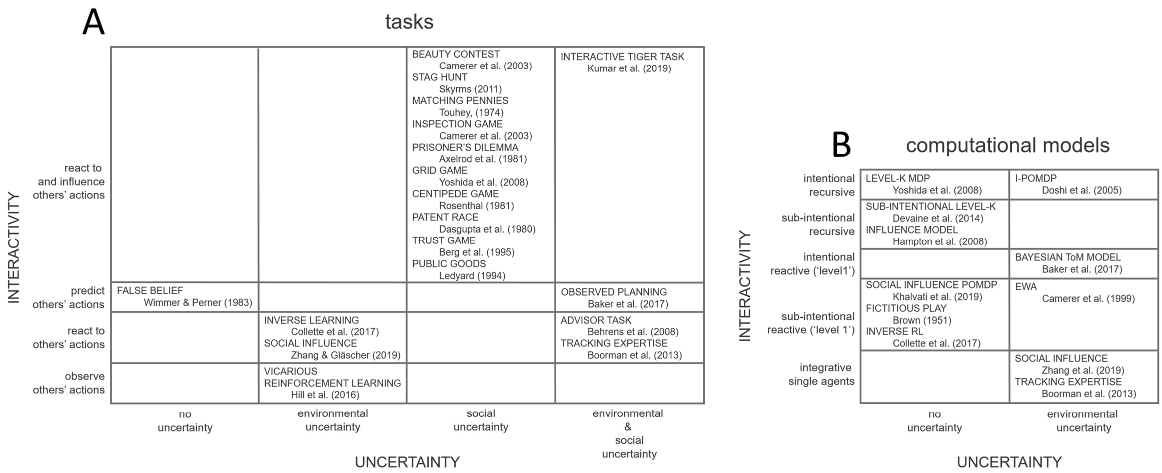

***Figure 4*| Localizing tasks and models with respect to our 2-dimensional classification space**
We characterize tasks and models with respect to two dimensions: interactivity and uncertainty.
(A) Uncertainty with respect to tasks is further split into social and environmental uncertainty. Social uncertainty refers to ambiguity about others' internal variables such as their individual preferences or level of cognitive sophistication, and their beliefs, values and motivations. Environmental uncertainty indicates that participants and observed agents are faced with a noisy and only partially predictable surrounding context and need to infer its current state on each trial. Note, that the false belief task includes neither social nor environmental uncertainty. However, this and other tasks include divergent belief states. This component is not captured in this figure. Interactivity is graded as follows: (1) mere observation of others actions without predicting or reacting to them but using the information their actions reveal about the environment, (2) reacting to others' choices by integrating them into one's own decision process but without taking others' perspectives, (3) predicting others' behavior from their frames of reference, and finally (4) predicting others by taking their perspectives while at the same time integrating this into one's own frame of reference to react to and influence others' behavior.
(B) Uncertainty with respect to models denotes whether they include parameters that model agents' learning about environmental and social uncertainty or include no learning. Interactivity indicates how well models parameterize representations of other agents. This ranges from (1) single agent models with no explicit representation of the other but integrating information derived from others' actions, (2) sub-intentional or (3) intentional representations of others to predict future choices and optimization of own actions with respect to these predictions, to (4) sub-intentional and (5) intentional recursive models where predictions of others include predicted responses to one's own actions and optimization of one's own behavior to these nested predictions.

3. Models

To specify the hidden cognitive processes underlying overt behavior cognitive neuroscience leverages quantitative computational models that are fit to the participants' behavioral responses. From these models, latent variables capturing the significant computations are extracted and combined with neural data in model-based analyses approaches. Model-based analyses in cognitive neuroscience examine how particular cognitive operations are carried out at the neural level (Gläscher & O'Doherty, 2010). We review a number of computational models that have been used to elucidate social decision-making processes. As was the case when characterizing experimental approaches, we will use the dimensions of interactivity and uncertainty to characterize models. Interactivity in the context of formal models denotes the extent to which agential models capture the other agent's reasoning. Interactivity in models is subdivided into two additional sub-dimensions: (1) intentionality, or models that include an intentional model of others in contrast to models that include only the effects of others' actions as regularities in the environment, and (2) recursivity, or models that capture processes such as "Agent A thinks that agent B thinks that A thinks that XYZ is the case". Uncertainty in models refers to models' capacities to represent the underlying but sometimes only partially accessible states and dynamics of the environment.

3.1. Non-interactive models for decision making under uncertainty

The formal social decision frameworks included here are based on single-agent reinforcement learning (RL) models. To simplify the understanding of subsequently presented social decision frameworks, we briefly summarize single agent models. RL problems are typically modeled by the Markov Decision Process (MDPs) (Puterman, 1990). An MDP is defined by set of states $S = \{s^1, \ldots, s^k, \ldots, s^n\}$ representing the environment, a set of actions $A = \{a^1, \ldots, a^k, \ldots, a^n\}$ an agent can take, a reward function determining the reward based on states and actions $R(s_{t-1}, a_{t-1}, s_t)$, and a transition function $T = p(s_t | s_{t-1}, a_{t-1})$ determining the environmental dynamics. The transition function captures the probabilities of transitioning between states given specified actions. Different decision strategies in the different states of the environment provide varying rewards to the agent. The goal in RL is to take those actions that maximize the long-term expected future rewards (Sutton and Barto, 2012). This can be achieved via choice-heuristics and learning the value of chosen (and sometimes unchosen) actions without explicitly representing the structure of the environment. This occurs by mapping actions directly to rewards in so-called "model-free learning". Alternatively, agents might develop "model-based learning" via a sophisticated representation of the environment, which means representing the transition probabilities between states, thereby allowing for flexible goal directed decision making. Evidence for both model-free and model-based learning has been found in humans and other organisms (Daw & Dayan, 2014; Daw, Niv, & Dayan, 2005; Gläscher, Daw, Dayan, & O'Doherty, 2010; Wan Lee, Shimojo, & O'Doherty, 2014).

3.1.1. Model-free single agent decision models

Model-free learning and decision making can be captured by temporal difference (TD) learning (Sutton and Barto, 2012). In TD, an agent learns solely by experience without knowing and representing the dynamics of the environment. At each time step t the value $V(s_t)$ of taking a strategy in state $s_t \in S$ of the environment is updated based on the obtained reward:

$$V(s_t) = V(s_t) + \alpha[r_{t+1} + \gamma V(s_{t+1}) - V(s_t)]$$

with α being a learning rate that weights the influence of new observations, γ being the discount factor regulating the effect of future values. Although lacking an explicit representation of the surrounding dynamics, dynamic value updating provides a basic representation of environmental uncertainty.

Basic model-free, single-agent RL algorithms can capture vicarious learning mechanisms elicited in observational learning scenarios in which individuals do not act themselves, but learn their own action values by observing the actions of others as well as the outcome of those actions (Burke, Tobler, Baddeley, & Schultz, 2010; Hill et al., 2016). More complex observational learning problems require extensions of classic single-agent RL models. To formalize indirect learning about the environment that occurs only by observing others whose preferences are known to the observer but whose action-outcomes are hidden (Figure 1E), Collette and colleagues (2017) used an inverse RL framework. Inverse RL distinguishes from classic RL in that instead of trying to find an optimal strategy with respect to a given reward function, it aims at inferring the reward function that best explains an agent's behavior (Arora and Doshi, 2018; Ng and Russell, 2000). Collette et al. (2017) used inverse RL to capture how humans make sense of the world by observing other agents' behavior. Their computational model recovered the underlying reward distribution that best explained the observed agents' actions, thereby capturing participants' learning about the environment only by observing behavior.

3.1.2. Adapted model-free single-agent models

A different adaptation of single-agent RL was found to best capture the tracking of others' expertise (Boorman et al., 2013) (Figure 1D). Boorman and colleagues combined classic learning about the values of strategies with learning about the quality of others' behavior (i.e., the correctness of others' actions). Observed agents' expertise was operationally defined as the probability of them making correct choices and was modeled in two sequential learning steps. First, the match between the observed agent's choice and one's own action-value estimate was assessed. Second, the expertise estimate was updated based on whether the observed action was or was not correct. The authors noted that the first updating step is suboptimal with respect to rational choice theory but was required to adequately model participants' choices. This suggests that instead of representing other agents as performing at a constant rate throughout, participants represented the agents that they observed as learning about the value of assets in a way that was similar to their own learning. Interestingly, this suboptimal model best explained participants' choices both when they received instructions that the "other" was a person or a computer.

A third variation of single agent RL was presented in a recent attempt to model social influence of group decisions on individual decision by Zhang & Gläscher (2019). In a two-phase group decision task, participants could adapt their own choice after the decisions of the other group members had been revealed (Figure 1C). Zhang & Gläscher developed a computational model that combines learning from one's own experiences (via a classic RL approach) with learning from other players. They did this by computing a value based on the recent reward history of the others. These two value signals were weighted into a single choice value determining the first decision. The model then predicted switch or stay after the disclosure of group behavior by incorporating parameters for the difficulty of the first decision and for the coherence of the group's decision.

All three variations of model-free, single-agent RL presented above capture individuals' learning about an uncertain environment when that learning is based on one's own action-outcome associations and/or observing others' behaviors in the world. In the inverse RL model by Collette et al. (2017) this is achieved by learning "through the eyes" of an observed agent. Although this inverse RL model does not include an interactive component and no explicit representation of others, it does include an intentional model of the other's decision-making processes. The observed agent's learning is explicitly represented to infer the hidden aspects of the world. In contrast, neither the social influence model nor the expertise tracking model explicitly represents others' learning or makes predictions about the choices of the agent being observed. These models represent others' task capacities irrespective of the decision processes that lead up to their choices. So, while these models capture learning about an uncertain surrounding from the agent's own perspective, they do not represent others' decision-making processes.

3.1.3. Model-based single agent models

Model-based action planning relies on an explicit representation of the environmental dynamics captured by the transition function T. Representing these dynamics allows for flexible decision-making. However, planning an optimal path through a given environment requires a representation of the current state. This is complicated when agents cannot directly observe the current state but only receive incomplete information about it. Under these conditions, percepts are partial and/or information is ambiguous. Formally, decision making under state uncertainty is captured by Partially Observable Markov Decision Processes (POMDPs) (Kaelbling et al., 1998). In a POMDP, information about the state at any given time is defined as a set of observations $\Omega = \{o^1, \ldots, o^k, \ldots, o^n\}$ an agent can make. Actions taken at a given time step a_t and the state s_{t+1} resulting from those actions define observation probabilities $O(o_t|s_t, a_{t-1})$. To deal with state uncertainty an agent integrates observations to form a belief b_t about the possible states of the environment. Agents' beliefs then dynamically update given the agents' observations and prior beliefs, using a Bayesian estimation function:

$$b_t(s_t) = \beta O(o_t|s_t, a_{t-1}) \sum_{s_{t-1}} T(s_t|s_{t-1}, a_{t-1}) b_{t-1}(s_{t-1}),$$

where

$$\beta = \frac{1}{\Pr(o_t|b_{t-1}, a_{t-1})} = \sum_{s_t} O(o_t|s_t, a_{t-1}) \sum_{s_{t-1}} T(s_t|s_{t-1}, a_{t-1}) b_{t-1}(s_{t-1})$$

is a normalizing constant.

To capture the core features of ToM, which are representations of others' belief and motivational states, Baker et al. (2017) used an extended POMDP model ("Bayesian ToM model"). This model computes a joint posterior probability representing an observer's beliefs about an observed agent's possible beliefs. The overall likelihood is factorized into a model for the observed agent's beliefs and a model of the agent's planning process based on beliefs and desires. The Bayesian ToM model formalizes detached non-interactive representations of an individual's intentional representation. These include another person's beliefs about an uncertain environment based on the observed agent's imperfect perceptual capabilities, along with the agent's subjective preferences inferred from observed behavior. Thereby the model captures intentional decision making from another individual's perspective. However, it does not formalize how this process is integrated with one's own beliefs and desires and the choices that one makes as a result of these beliefs and desires.

Both, model-free and model-based single-agent RL frameworks capture environmental uncertainty. Additionally, extensions of these models use inverse RL and the Bayesian ToM model to formally represent others' intentional learning processes from a detached observational perspective. Active engagement and interaction are not formal features of these models.

3.2. Interactive decision models without environmental uncertainty

3.2.1. Recursive models

When multiple agents engage in interactions where their joint actions are relevant to one another, such that their individual rewards depend on others' behavior, reasoning can become recursive in the form of "I think, that you think, that I think, ..." and so on. Formally, this sort of thinking is captured by what is called a "level-k framework" (Stahl, 1993) and cognitive hierarchy theory (Camerer et al., 2004). The depth of reasoning an agent engages in is referred to as its *level* with the parameter k determining the sophistication or depth of reasoning. The concept is well illustrated by the beauty contest game described before (Figure 2A). Level-k frameworks are defined in a bottom-up fashion starting with a base level 0 agent model. A level-0 agent does not reason about others. It might completely discard any information about the other's behavior and treat it as environmental noise, or assume the co-player is acting according to a hidden distribution (Coricelli & Nagel, 2009; Devaine et al., 2014; Gmytrasiewicz & Doshi, 2005; Yoshida et al., 2008). A level-k agent represents the other agent at level k-1. That is, a level-1 agent A1 represents the other agent A2 at level-0 (i.e. as having no ToM). A level-2 agent A1 represents the other agent A2 as a level-1 agent. This means that from the perspective of A1, A2 represents A1 as a level 0 agent. This illustrates that by definition, k-level models represent other agents as ill-informed about the ToM level of the primary agent. The recursion in k-level frameworks theoretically extends ad infinitum, but typically most human choices do not require modelling agents beyond level 3 (Camerer et al., 2015). The specifics of level-0 models and consequently all higher-level definitions are determined by the decision problem and the underlying basic decision model.

The advance of level-k frameworks over other models reviewed before is the capability to formally model agents' representations of others as interactive agents themselves, i.e., as agents that respond to others' behavior and that have representations of one's own ToM. This allows capturing real interactivity and interwoven information processes as "model within a model". Second, specifying different "base models" as level-0 models allows the capture of a variety of social reasoning processes; from representing others' as following simple sub-intentional strategies to representing others as fully intentional, goal directed agents. Here, we provide an overview over the most prominent of these models and we highlight their properties with respect to our two defining dimensions and elucidate their implications for ToM research.

3.2.2. Level 1 and level 2 models

Although not actually considered a level-k model, fictitious play is a basic framework for interactive decision making (Brown, 1951). In a fictitious play model, an agent observes the history of a co-player's actions and forms expectations about the co-player's future actions based on the frequency of past choice. In essence, the agent tracks the most frequently selected action in the past. With respect to these predictions the agent then chooses the action maximizing its own rewards. The choice history is dynamically tracked via a simple updating rule essentially counting the co-player's frequency of taking any of the available actions. Fictitious play represents co-players' behaviors via a very basic choice heuristic (namely, "what the other has frequently chosen previously, will likely be his choice in the future") and represents others as sub-intentional level-0 agents. Moving up one step in the reasoning hierarchy, Hampton et al. (2008) introduced an influence model. Instead of directly predicting an opponent's behavior from the choice history, the model computes the influence that one's own behavior has on the opponent, assuming that the opponent uses fictitious play. The agent then optimizes its own choices with respect to this prediction.

Capturing interactive behavior in a public goods game (Figure 2B) Khalvati et al. (2019) utilized a level-1 sub-intentional social influence variant of the POMDP framework. In their version of the task, all other players in the group were displayed as identical avatars, thus rendering the social interaction anonymous (i.e., the identities of the other individual players could not be inferred). As a consequence, a computational model for the task is best when it models only the mean contribution probability of the entire group and not the individual tendency of each member of the group to contribute to the public good. In their implementation of a POMDP model applied to the public goods task, Khalvati and colleagues used a beta distribution that is updated on every trial to represent the participant's belief about the overall group contribution to the public good. Effectively, the model represents each individual player as a level-0 agent that chooses to contribute with the probability defined by the belief distribution. This belief is updated over time using a Bayesian learning rule and participants' observations, capturing participants' learning about the groups' behavior. The decision processes that influence whether the group members contribute or free-ride, as well as any effects of one's own choices on those decision processes, are neglected. The resulting model does not incorporate ToM because it represents the individual group members as sub-intentional decision makers whose actions are captured by simple action probabilities.

3.2.3. Fully recursive modeling

A full level-k framework for strategic decisions was introduced by Devaine and colleagues (Devaine et al., 2014) to investigate different decision strategies when playing against a supposedly human opponent vs. a random computer agent in a simple matching pennies task. In their model, level-0 opponents' choice probabilities p_y are assumed to follow a time-varying hidden distribution x^t. Observing actions provides information about the mean μ_t^0 of the underlying distribution. Similar to a prediction error, μ_t^0 is updated at each time point with the difference in observed and expected action:

$$\mu_t^0 = \mu_{t-1}^0 + \sum_t (a_y^t - s(\mu_{t-1}^0))$$

with s denoting a sigmoid choice function. The opponent's decision probability $p_y^{t+1}(a)$ is then computed with a sigmoid decision function based on μ_t^0 and an unknown volatility σ^0 and choice temperature β^0. Based on this action probability for the opponent, the level-0 agent chooses its own action such that it maximizes the individual expected value. In a recursive fashion, predictions about the opponents' choices for a level-1 agent and higher are built up from this level-0 decision rule. The agential model needs to select both the hidden variables governing action probabilities and the sophistication (k-level) of the agent's opponent. This results in a posterior distribution over the opponent's k-levels and the action probability variables. These estimates in a fully recursive framework allow examining how participants represent others' cognitive abilities and their beliefs about how they are themselves represented by others.

Irrespective of sophistication, all models considered in this section are based on sub-intentional level-0 models. That means that at the lowest level, others are not represented as agents with desires, beliefs or intents but as agents whose actions follow simple hidden distributions. Thereby, at the lowest level these models capture others' actions as non-intentionally rooted information that is integrated into one's own decision processes. However, the level-k framework by Devaine et al. (2014) and Hampton et al.'s influence model (2008) include higher-level representations. These comprise intentional representations of others, although they are still based on a sub-intentional level-0 model. Further, neither basic fictitious play, the influence model, nor the full level-k reasoning model include a representation of the environment. Therefore, they are only applicable to static strategic games taking place in a stable environment such as matching pennies or the inspection game (Figure 2C) (Camerer, 2003; Devaine et al., 2014; Hampton et al., 2008).

Yoshida and colleagues (2008) capture strategic decision making in the spatial stag-hunt game (Figure 2F). The spatial variation of this simple coordination game requires long term action planning with respect to the (fully observable) surrounding and the co-player's future actions. To integrate both agents' acting in the environment they extend a basic MDP model by defining the state space as the product of both agents' admissible states. Consequently, rewards are also defined on the joint state space and the other's actions are a function solely of the other's private value function. Based on the interacting agents' estimates of their respective goals, the models predict cooperative, coordinated or individualized behavior. In the next step, optimal strategies with different levels of recursion were computed in the extended multi-agent MDP framework. The resulting k-level MDP model can capture model-based goal directed agents recursively integrating other intentional agents' behavior into their decision process. However, the underlying MDP is fixed and environmental properties are not dynamically learned by the agents. Hence, the k-level MDP represents no learning about the environment.

3.3. Interactive models for decision making under uncertainty

In this final model characterization, we introduce two computational frameworks that extend reinforcement learning to the multi-agent domain, allowing us to formalize interactive social decision making under social and environmental uncertainty. Although both models can be applied to larger multi-agent scenarios, for the sake of simplicity and comprehensibility we only describe a two-agent implementation.

3.3.1. Non-hierarchical modeling

Experience weighted attraction (EWA) (Camerer & Ho, 1999) combines simple model-free RL and belief-based learning in a continuously weighted fashion. Simple RL estimates the values of one's own actions given the current rewards and updates these value representations via an experienced-based reward prediction error. It does not explicitly take the actions of the other players into account. Belief-based learning (essentially fictitious play) estimates the probability distributions by which the other agent chooses its actions, and then adapts the chosen action accordingly. These two individual forms of learning are implemented in EWA by specific parameter settings (Camerer & Ho, 1999). However, the power of EWA lies in the continuous weighting of these two forms of learning by a trade-off parameter δ (see below). Thereby, EWA combines a representation of the environment, which is learning the value of own actions given the current state of affairs, with a simple (i.e., sub-intentional) model of others' actions. The value of all available actions is updated according to a learning rule that combines several variables. The first variable $N(t)$ is equivalent to previous (pre-measurement) experience and updated according to the following rule:

$$N(t) = \rho N(t-1) + 1.$$

The parameter ρ is a depreciation term that reflects how fast new reward associations can override prior experience. EWA updates the value of *all* available actions, but the update distinguishes between the chosen ($a_x^k = a_x(t)$) and the non-chosen actions ($a_x^k \neq a_x(t)$), while holding the action of the other player ($a_y^j(t)$) constant.

$$V_x^k(t) = \begin{cases} \dfrac{\varphi N(t-1)V_x^k(t-1) + \delta \pi_x(a_x^k, a_y^j(t))}{N(t)}, & if a_x^k \neq a_x(t) \\ \dfrac{\varphi N(t-1)V_x^k(t-1) + \pi_x(a_x^k, a_y^j(t))}{N(t)}, & if a_x^k = a_x(t) \end{cases}$$

Thus, in case of the actual action ($a_x^k = a_x(t)$), the value of that action is updated with the full joint reward $\pi(a_x(t), a_y(t))$ whereas the unchosen joint action $a_x^k \neq a_x(t)$ is updated with the delta-weighted joint reward $\delta * \pi(a_x(t), a_y(t))$. It is one of the key insights of EWA (detailed in Camerer & Ho, 1999) that this weighted updating of unchosen action approximates belief-based learning. EWA has been successfully used in several behavioral economics experiments (see Camerer et al., 2004) and in a decision neuroscience context for modeling choice behavior in patent race games (Zhu et al., 2019, 2011). However, it is a computational model that provides only a very basic, sub-intentional, non-recursive representation of others.

3.3.2. Fully recursive modeling

A combination of model-based RL and intentional recursive representations of others' decision processes is given by the Interactive Partially Observable Markov Decision Processes (I-POMDPs) (Gmytrasiewicz and Doshi, 2005). Recall that an (individual) POMDP agent x forms beliefs $b_x^t(s_t)$ about the physical states of the environment allowing it to plan actions in an uncertain, only partially accessible surrounding. I-POMDPs extend single-agent POMDPs to the interactive domain by combining the physical state space S_{phys} with intentional models of the other agent y (Θ_y) yielding an interactive state space $IS_x = S_{phys} \times \theta_y$. Consequently, an I-POMDP agent's beliefs are no longer over S but over IS: $\mathrm{b_x^t(is_t)}$. The key component of the model θ_y which agent x forms about the second agent y is that it includes y's beliefs $\mathrm{b_y^t}$. That means agent x's belief $b_x^t(is_t)$ is a probability distribution over the multidimensional space spanned by physical states and beliefs of y and hence captures x's belief about the state of the physical environment and y's belief. These aspects of the model are then the agent's knowledge about the world and its knowledge about another individual's intentional states. Agent y's beliefs are either over the physical states space $b_y^t(s)$ (essentially equal to a single agent POMDP) or these beliefs themselves can be over an interactive state space including models of x ($IS_y = S_{phys} \times \theta_x$). The first case, in which agent y forms beliefs over physical states only, results in a level-1 model for agent x. Agent x represents agent y as an intentional goal-directed level-0 agent that acts to maximize its reward in the world but does not represent or react to other agents in the surrounding. In the latter case, in which agent y forms beliefs about x's beliefs, agent x represents y as reasoning about himself. That results in a level-2 model for agent x. Theoretically, as discussed before, this recursion could go on and beliefs could be nested infinitely yielding higher-level agent x models. However, as for simpler level-k frameworks, it is reasonable to assume bounded rationality. Dynamic belief updating in the I-POMDP framework follows the same basic Bayesian learning rule as the POMDP update, however it is extended to include an update of the other's belief (for details see Gmytrasiewicz & Doshi, 2005). This requires solving models of y in bottom-up manner, essentially simulating y's learning and decision process. Environmental uncertainty and uncertainty regarding the other's model of oneself hamper this process and make model parametrization and selection challenging.

The I-POMDP framework has been used to model behavior and reasoning processes in a multi-round trust task (Figure 2G) (Hula et al., 2018, 2015) and the centipede game (Figure 2I) (Doshi, Qu, Goodie, & Young, 2012, Doshi, Qu, & Goodie, 2014). In its application to the trust task, the I-POMDP allowed formalizing players' risk aversion and guilt as well as the recursive representations of the co-player's risk and guilt parameters. These parameters define the rate of cooperation and the breakdown and reestablishment of cooperation between players over the course of multiple interactions (Hula et al., 2018). Using I-POMDP models to capture decision processes in a centipede game showed that participants mostly adopt decision strategies of level-1 or level-2 and that the depth of recursive reasoning increases in more competitive scenarios (Doshi et al., 2012).

While EWA formalizes an agent's own knowledge state it cannot model others' beliefs. Co-player's actions are included by tracking the frequency of forgone choices. Thereby, the framework fails to capture the intentional representations of others' knowledge states and resulting decision processes. This is different from the approach taken by the I-POMDP framework. By recursively meshing POMDP models, I-POMDP models formalize high level ToM processes. Others are represented as intentional goal-directed agents whose beliefs are dynamically updated as information about the environment unfolds. Further, via recursion, others' intentional representations of one's own beliefs, values and motivations can be formalized and so quantitatively represented. The I-POMDP framework is very well suited to model complex ToM processes in a range of applications.

4. Neural responses

In previous sections we addressed the primary focus of the typological proposal in this paper, which draws on a detailed description of social decision-making tasks and their computational models. We were able to do so, because of the wealth and amount of behavioral studies that have used these tasks and that have analyzed the data using computational models we summarized. The aim of this section is to characterize how the elicitation of neural responses may correspond to our typology involving uncertainty and interactivity in cognitive tasks investigating ToM. Therefore, our focus is different from those of previous reviews (Amodio & Frith, 2006; Frith & Frith, 2006; Mitchell, 2009; Saxe, 2006) and meta-analyses (Schurz et al., 2014; Van Overwalle and Baetens, 2009), which parcellated the available studies based on different ToM tasks (e.g. false belief tasks, trait judgments, social animations, the mind in the eye task, strategic games). However, only few neuroimaging studies can be described in terms of interactivity and uncertainty, and even fewer have used model-based fMRI analyses (Gläscher & O'Doherty, 2010), which represents the current state-of-the-art for relating computational models as described above directly to neuroimaging data. In addition, previous work has presented analyses of the neuroimaging data (e.g. the specific model-based contrasts) that does not necessarily address the dimensions of interactivity and uncertainty that define the typology of this review. Therefore, in this section we describe the neural responses in terms of (a) representing others' beliefs and intentions, and (b) recursive ToM.

Although early neuroimaging studies have employed famous interactive decision-making tasks like the prisoner's dilemma and the trust game, the analyses have generally focused on the comparison of cooperative vs. competitive behavior (Rilling et al., 2002), the faces of cooperative vs. competitive opponents (Singer et al., 2004), the reputation to cooperate (Phan et al., 2010), or simply on good vs. bad outcomes (Delgado et al., 2005). Common to these findings is a robust activation of the striatum (ventral and dorsal striatum, putamen, and caudate head) and vmPFC (Li et al., 2009) when contrasting cooperative with competitive behavior by the other player. This resonates with many (single-agent) reward learning studies that report reward-related activation in this region (Bartra et al., 2013), including reward prediction errors (RPEs) (Garrison et al., 2013) suggesting that pro-social interactions may act as a social reward.

A substantial number of neuroimaging studies that have used false belief tasks to investigate the representation of others' beliefs have reported activations in bilateral TPJ, mPFC, and precuneus (see Schurz, 2014 for a meta-analysis). Several neuroimaging studies have also investigated representations of others with model-based approaches. A common finding among these studies is the involvement of the medial and dorsomedial prefrontal cortex in these representations, similar to reports from earlier reviews and meta-analyses. For instance, Behrens et al. (2008) using the original advisor task reported a social prediction error in the dmPFC and rTPJ/pSTS, whereas an RPE correlated with BOLD activity in the ventral striatum, consistent with the findings mentioned. Moreover, Collette et al. (2017) reported an activation of mPFC correlating with simulated values of an observed player. In that study, rTPJ was associated with a belief entropy signal which was related to the uncertainty of current beliefs. In contrast, in the spatial stag hunt game (Yoshida et al., 2010) mPFC activity correlated with a belief entropy signal, coding the uncertainty about the beliefs of the other player. Similarly, mPFC also correlated with the belief estimates of the observed person's ability in the expertise learning task (Boorman et al., 2013), whereas rTPJ was linked to a belief updating signal. In a similar vein, mPFC was associated with the difference in log-likelihood between the influence model and a simple fictitious play model in the inspection game (Hampton et al., 2008), suggesting that it was related to level-2 beliefs about influences of the opponent's choice. Zhang & Gläscher (2019) reported that activity in bilateral TPJ/pSTS correlated with the number of other players with opposing decisions. In that report, vmPFC was related to the expected value learned from one's own experience compared to the value learned from the other players' recent reward history. Despite the differences in these tasks and where they fall on our interactivity dimension (see Figure 4A), the commonality of these neuroimaging findings suggests that bilateral vmPFC and dmPFC are often recruited during the representation another person's beliefs and abilities. The computational role of the TPJ - though clearly and robustly involved in many social decision-making paradigms - remains elusive. This suggests that the information processing contributions of this region are multi-purpose and that networks within the region can be recruited to perform different computations in different experimental contexts.

Another region that is often related to representing aspects of a social partner and interactions with them is the anterior cingulate cortex (ACC). For instance, in comparing the trust game with a control game, the ACC is related to trust decisions (Krueger et al., 2007), whereas the septal area and the ventral tegmental area are more specifically related to building and maintaining trust. During a vicarious RL task involving students who learn and an all-knowing teacher (Apps and Ramnani, 2014), the activity in the ACC reflects prediction errors signals for the teacher's simulated values of the students' value estimate. Similarly, in the expertise tracking task the ACC was involved in computing a belief updating signal in the form of an "ability prediction error". Moreover, Zhu et al. (2011) reported belief prediction errors about the opponent's actions in the rostral (perigenual) part of the ACC. In the original volatility learning task (Behrens et al., 2007) as well as in its social variant, the advisor task (Behrens, Hunt, Woolrich, & Rushworth, 2008), the ACC correlates with a model-derived volatility signal of the environment or of the social partner. This volatility signal in turn influences the first-order learning rates that update reward expectations. In the social influence task by Zhang & Gläscher (2019), the ACC represents the value signal computed from other players' recent reward histories. In summary, similar to the rTPJ, the networks in the ACC often, but not exclusively, show activation patterns that covary with error signals that index violations of expectations about the environment or of social partners. This pattern of error-related activation in the ACC is consistent with its well-documented role in error monitoring (Holroyd and Coles, 2002).

Few studies have directly investigated recursive ToM and the different levels of sophistication that make up the highest level in our typology of social decision-making tasks. Bhatt and Camerer (2005) used a series of matrix games that are "dominance-solvable" meaning that through iterated reasoning, non-optimal strategies can be eliminated as one identifies the equilibrium strategy. During the experiments they asked the participants to simply make their own choice (level-0), estimate what the other player is going to choose (level-1 beliefs), and guess what the other player thinks they will choose (level-2 beliefs). They identified the left anterior insula and the right inferior frontal gyrus when contrasting BOLD responses from games with level-2 vs. level-1 questions. In contrast, rACC, posterior cingulate cortex (PCC), and dlPFC showed stronger BOLD responses when making choices compared to level-1 beliefs.

Employing model-based fMRI analysis, Yoshida (2010) used the spatial stag hunt game and their previously developed computational model (Yoshida, 2008), to identify neural correlates of belief uncertainty (entropy) about the computer agent's strategies. In that analysis, the trial-by-trial estimate of the agent's sophistication level correlated with activation in the superior parietal lobule, the frontal eye fields, and the dlPFC (albeit in much more dorsal than reported in the Bhatt & Camerer study).

While in the stag hunt game estimating the level of reasoning is done by comparing model predictions at different levels of reasoning, the beauty contest game offers a more direct estimation. The choices made by the participants directly reflect how far participants iterated their own strategy with those of the entire group. Coricelli and Nagel (2009) used a version of this game adapted to the fMRI environment and instructed participants to make choices. Participants played against human opponents or against a computer simulation of group decisions. Using the cognitive hierarchy model (Camerer et al., 2004) to analyze the behavioral data, they observed that most participants showed levels of ToM between level-1 and level-3. Based on the distance between the data and model predictions, they classified participants into high and low levels of reasoning. These subgroups exhibited activations in the rACC for low, and in the vmPFC and mPFC for high level reasoning.

In the previous two studies, the participants' choices revealed their reasoning level in response to partner or group decisions. However, the modeling in these studies did not take the influence that the participant might exert on the other players into account. The inspection game (Hampton et al., 2008; Hill et al., 2017) addresses this aspect of interactive reasoning. Although the modeling does not explicitly refer to the level of reasoning, the authors contrast different computational models (Reinforcement Learning, Fictitious Play, Influence Model) that correspond to different levels of recursivity in ToM. In particular, the Influence Model captures the influence that the participants exert on their opponents thus elevating the reasoning process to level-2. Expected value signals derived from the influence model correlated with brain activity in vmPFC more strongly than those derived from other, less sophisticated models. Belief updating signals from the influence model also correlated with activation of the rTPJ. The conviction that one was influencing one's opponent, measured as the difference in the log-likelihood of the influence and fictitious play models, showed a robust activation of mPFC.

Lesion-Deficit Analyses (LDA) have also attempted to segregate the high-level, inference-based ToM network, which includes the regions already discussed above, from lower-level, perception-based networks. The latter are sometimes called simulation networks and include the anterior intraparietal sulcus (aIPS) and premotor cortex (PMC) (Van Overwalle and Baetens, 2009). In a group of patients with a rare form of glioma that migrates along large associative fiber tracts, Herbet et al. (2014) were able to link impairment in high-level ToM to surgical disruptions of the arcuate fasciculus, whereas lower-level ToM was associated with disruptions of the cingulum. This emphasizes that the functional ToM network is built on structural connectivity and can be disrupted by severing significant white matter connections. These results for high-level mentalizing were later confirmed in additional glioma patients, which also highlighted the superior longitudinal fasciculus and the fronto-striatal tract, whereas lower-level and face-based mentalizing was also associated with the OFC and the uncinate fasciculus (Nakajima et al., 2018a, 2018b). Furthermore, a recent functional connectivity study (Fishman et al., 2014) found that in contrast to commonly held beliefs, persons with AS compared to NT controls exhibit increased and not decreased connectivity between the mentalizing and the simulation networks, giving rise to the intriguing hypothesis that persons with AS may suffer from overconnectivity and – as a result – a diminished functional segregation between these two networks. However, this finding could only be demonstrated in 15 tightly matched pairs of participants and thus it awaits replication in a large sample.

In conclusion, while a robust ToM network including of TPJ, mPFC/rACC, precuneus, and vmPFC, and sometimes also the dorsal ACC, is consistently recruited in various different ToM tasks, the specific roles of each of these network nodes remains multidimensional and requires further specification. Others have attempted to parse the heterogeneity of findings in the mentalizing network in terms of self-referencing and other-referencing information processing (Joiner et al., 2017). However, they also conclude that different computational signals appear to be represented in the same brain regions for different tasks. The evidence from the few model-based fMRI studies reviewed here also suggests a dynamic recruitment of these areas when accomplishing related, but distinct, tasks that involve different degrees of interactivity and uncertainty. The array of different tasks and the lack of attention to how these tasks differ has likely contributed to the heterogeneity of interpretations thus far and has contributed to lack of clarity regarding the computational roles of the nodes in the ToM network. It is our hope that with additional model-based neuroimaging studies in this field, possibly designed along our axes of interactivity and uncertainty, a more precise characterization of the computations will emerge.

5. Conclusion

When aiming at examining human ToM capacities, several important aspects need to be considered. First, one should be aware that Theory of Mind is a highly inclusive concept that implies a variety of cognitive sub-functions including emotional processes, motivational and goal-oriented valuational processes, and functions associated with belief and knowledge (Schaafsma et al., 2015). Furthermore, as shown in the first section of this review, even within one “subsection” of these functions the cognitive processes that are likely elicited differ strongly depending on the specifics of the social situation. We argued that the two dimensions of uncertainty and interactivity can provide an effective typology of tasks for understanding the potential for eliciting the varying levels of ToM in social decision making. First, we proposed that uneven distribution of information about the environment among agents and increased uncertainty about the environment likely elicits representations of others’ belief states. Second, we suggested that uncertainty about others’ motivational traits and dynamically changing states creates an increased functional relevance for representing others’ motivational traits and states. Finally, we propose that behavioral relevance and the interdependence of individuals’ actions determines the frame of reference. In tasks that do not directly link one’s own successful outcome to one’s representation of others’ beliefs and motivations, the outcomes of others’ choices can be used as an additional source of information to guide one's own actions. However, in such task situations there is little to no incentive to engage in representing the intentionality of those others. Hence, integration of this information into individuals' own frame of reference without taking the others' perspective is sufficient. Predictions under asymmetric belief states and uncertainty, however, require taking the others' intentional perspectives into account via some level of representation. Lastly, we suggested that true action-interdependence under uncertainty requires the integration of one's own and others' intentional perspectives and decision processes, best allowing for scientific inquiry into high level ToM. Consequently, tasks aimed at investigating ToM processes in their full richness should be designed with both of these dimensions in mind. Further, for a more complete and ecological validity picture of ToM processes, tasks should explore rich environments and face-to-face interactions.

In the second section of this review we characterized formal computational frameworks and identified models' varying capacities to map uncertainty and to integrate multiple agents' beliefs, motivations and decision processes. Computational models provide quantitative testable descriptions of hidden cognitive functions and their putative parameters. Applying models that capture uncertainty and interactivity might allow us to disentangle the multitude of sub-processes of ToM. We argued that different tasks elicit different grades of ToM. Validating these claims requires testing a variety of social decision models on these tasks to objectively characterize if representations of others differ in the various scenarios. Thereby, we might gain a more complete and structured understanding of the cognitive processes underlying social decisions and would be able to examine the interplay of underlying neural systems in more detail.

Here we focused on the importance of interactivity with respect to the interaction of one's own and others' referential cognitive processes. We note that a call for strong interactivity has previously been made by proponents of "second-person neuroscience". Advocates of this view argue that cognition during interaction differs fundamentally from observational scenarios. They argue that not only is recursive thinking elicited only during interaction, but they also stress qualitative components like the feeling of engagement with or connection to others that come into play during interaction (Redcay and Schilbach, 2019; Schilbach et al., 2013). Further, they emphasize the importance of "multi-brain" or "hyperscanning" studies during which the neural activity of multiple interacting agents is recorded. Hyperscanning recordings have successfully been conducted using fMRI and EEG and have revealed specific synchronizations between the neural activity of interacting partners during abstract communication and motion matching tasks (Dumas et al., 2010; Stolk et al., 2015).

In line with these views, we argue for investigating ToM in rich interactive contexts under environmental and social uncertainty, while simultaneously recording neural activity from all interacting individuals. Such designs will provide measures that enable the discovery and parameter estimation of accurate models of the neural coding of ToM in its full complexity.

Acknowledgements

This review has benefited greatly from the comments of two reviewers. Further, we are grateful for helpful conversations with Martin Hebart, Christoph Korn, Yuqing Lei, Shannon Klotz, Corinne Donnay, Gregory Peterson, Robert Roberts, Jonathan Daume, and members of the L'Arche and Homeboy Industries communities. PD, SSK, MS and JG were funded by a Collaborative Research in Computational Neuroscience grant awarded jointly by the German Ministry of Education and Research (BMBF, 01GQ1603) and the United States National

Science Foundation (NSF, 1608278). JG and TR were supported by the Collaborative Research Center TRR 169 "Crossmodal Learning" funded by the German Research Foundation (DFG) and the National Science Foundation of China (NSFC). MS gratefully acknowledges funding from the John Templeton Foundation (Grant 21338) and the Templeton Religion Trust and the Self, Motivation, and Virtue Project. All authors declare no conflict of interest.

References

Amodio, D.M., Frith, C.D., 2006. Meeting of minds : the medial frontal cortex and social cognition. Nat. Rev. Neurosci. 268–277. https://doi.org/10.1038/nrn1884

Apps, M.A.J., Ramnani, N., 2014. The Anterior Cingulate Gyrus Signals the Net Value of Others' Rewards. J. Neurosci. 34, 6190–6200. https://doi.org/10.1523/JNEUROSCI.2701-13.2014

Arora, S., Doshi, P., 2018. A Survey of Inverse Reinforcement Learning: Challenges, Methods and Progress.

Assaf, M., Hyatt, C.J., Wong, C.G., Johnson, M.R., Schultz, R.T., Hendler, T., Pearlson, G.D., 2013. Mentalizing and motivation neural function during social interactions in autism spectrum disorders. NeuroImage Clin. 3, 321–331. https://doi.org/10.1016/j.nicl.2013.09.005

Axelrod, R., Hamilton, W., 1981. The evolution of cooperation. Science (4489). 211, 1390–1396. https://doi.org/10.1126/science.7466396

Baillargeon, R., Scott, R.M., He, Z., 2010. False-belief understanding in infants. Trends Cogn. Sci. 14, 110–118. https://doi.org/10.1016/j.tics.2009.12.006

Baker, C.L., Jara-Ettinger, J., Saxe, R., Tenenbaum, J.B., 2017. Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. Nat. Hum. Behav. 1, 1–10. https://doi.org/10.1038/s41562-017-0064

Baron-Cohen, S., 1988. Social and pragmatic deficits in autism: Cognitive or affective? J. Autism Dev. Disord. 18, 379–402. https://doi.org/10.1007/BF02212194

Baron-Cohen, S., Leslie, A.M., Frith, U., 1985. Does the autistic child have a "theory of mind" ? Cognition 21, 37–46. https://doi.org/10.1016/0010-0277(85)90022-8

Barrett, L., Henzi, P., 2005. The social nature of primate cognition. Proc. R. Soc. B Biol. Sci. 272, 1865–1875. https://doi.org/10.1098/rspb.2005.3200

Bartra, O., McGuire, J.T., Kable, J.W., 2013. The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. Neuroimage 76, 412–427. https://doi.org/10.1016/j.neuroimage.2013.02.063

Behrens, T.E.J., Hunt, L.T., Woolrich, M.W., Rushworth, M.F.S., 2008. Associative learning of social value. Nature 456, 245–249. https://doi.org/10.1038/nature07538

Behrens, T.E.J., Woolrich, M.W., Walton, M.E., Rushworth, M.F.S., 2007. Learning the value of information in an uncertain world. Nat. Neurosci. 10, 1214–1221. https://doi.org/10.1038/nn1954

Berg, J., Dickhaut, J., McCabe, K., 1995. Trust, reciprocity, and social history. Games Econ. Behav. https://doi.org/10.1006/game.1995.1027

Bhatt, M., Camerer, C.F., 2005. Self-referential thinking and equilibrium as states of mind in games: fMRI evidence. Games Econ. Behav. https://doi.org/10.1016/j.geb.2005.03.007

Boorman, E.D., O'Doherty, J.P., Adolphs, R., Rangel, A., O'Doherty, J.P., Adolphs, R., Rangel, A., 2013. The behavioral and neural mechanisms underlying the tracking of expertise. Neuron 80, 1558–1571. https://doi.org/10.1016/j.neuron.2013.10.024

Bora, E., Yucel, M., Pantelis, C., 2009. Theory of mind impairment in schizophrenia: Meta-analysis. Schizophr. Res. 109, 1–9. https://doi.org/10.1016/j.schres.2008.12.020

Brown, G., 1951. Iterative solution of games by fictitious play, in: Activity Analysis of Production and Allocation. Wiley, New York, pp. 374–376.

Bublatzky, F., Pittig, A., Schupp, H.T., Alpers, G.W., 2017. Face-to-face: Perceived personal relevance amplifies face processing. Soc. Cogn. Affect. Neurosci. 12, 811–822. https://doi.org/10.1093/scan/nsx001

Buitelaar, J.K., Van Der Wees, M., Swaab–Barneveld, H., Van Der Gaag, R.J., 1999. Theory of mind and emotion-recognition functioning in autistic spectrum disorders and in psychiatric control and normal children. Dev. Psychopathol. 11, 39–58. https://doi.org/10.1017/S0954579499001947

Burke, C.J., Tobler, P.N., Baddeley, M., Schultz, W., 2010. Neural mechanisms of observational learning. Proc. Natl. Acad. Sci. 107, 14431–14436. https://doi.org/10.1073/pnas.1003111107

Büyükboyacı, M., 2014. Risk attitudes and the stag-hunt game. Econ. Lett. 124, 323–325. https://doi.org/10.1016/j.econlet.2014.06.019

Byom, L.J., Mutlu, B., 2013. Theory of mind: Mechanisms, methods, and new directions. Front. Hum. Neurosci. 7, 1–12. https://doi.org/10.3389/fnhum.2013.00413

Byrne, R.W., 2018. Machiavellian intelligence retrospective. J. Comp. Psychol. 132, 432–436. https://doi.org/10.1037/com0000139

Call, J., Tomasello, M., 2008. Does the chimpanzee have a theory of mind? 30 years later. Trends Cogn. Sci. 12, 187–192. https://doi.org/10.1016/j.tics.2008.02.010

Call, J., Tomasello, M., 1999. A non-verbal false belief task: the performance of chimpanzees and human children. Child Dev. 70, 381–395.

Camerer, C., Ho, T.-H., 1999. Experience-Weightes Attraction Learning in Normal Form Games. Econometrica 67, 827–874.

Camerer, C., Ho, T., Chong, K., 2003. Models of Thinking, Learning, and Teaching in Games. Am. Econ. Rev. 93, 192–195. https://doi.org/10.1257/000282803321947038

Camerer, C.F., 2003. Behavioural studies of strategic thinking in games. Trends Cogn. Sci. 7, 225–231. https://doi.org/10.1016/S1364-6613(03)00094-9

Camerer, C.F., Ho, T.-H., Chong, J.-K., 2004. A Cognitive Hierarchy Model of Games. Q. J. Econ. 119, 861–898. https://doi.org/10.1162/0033553041502225

Camerer, C.F., Ho, T.H., Chong, J.K., 2015. A psychological approach to strategic thinking in games. Curr. Opin. Behav. Sci. 3, 157–162. https://doi.org/10.1016/j.cobeha.2015.04.005

Cañigueral, R., Hamilton, A.F. de C., 2019. Do Beliefs About Whether Others Can See Modulate Social Seeking in Autism? J. Autism Dev. Disord. 49, 335–348. https://doi.org/10.1007/s10803-018-3760-1

Charpentier, C.J., O'Doherty, J.P., 2018. The application of computational models to social neuroscience: promises and pitfalls. Soc. Neurosci. 13, 637–647. https://doi.org/10.1080/17470919.2018.1518834

Chaudhuri, A., 2011. Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. Exp. Econ. 14, 47–83. https://doi.org/10.1007/s10683-010-9257-1

Collette, S., Pauli, W.M., Bossaerts, P., O'Doherty, J., 2017. Neural computations underlying inverse reinforcement learning in the human brain. Elife 6, 1–20. https://doi.org/10.7554/eLife.29718

Coricelli, G., Nagel, R., 2009. Neural correlates of depth of strategic reasoning in medial prefrontal cortex. Proc. Natl. Acad. Sci. 106, 9163–9168. https://doi.org/10.1073/pnas.0807721106

D'Arc, B.F., Devaine, M., Daunizeau, J., 2018. A reverse Turing-test for predicting social deficits in people with Autism. bioRxiv 414540. https://doi.org/https://doi.org/10.1101/414540

Dasgupta, P., Stiglitz, J., 1980. Uncertainty, Industrial Structure, and the Speed of R&D. Bell J. Econ. 11, 1. https://doi.org/10.2307/3003398

Daw, N.D., Dayan, P., 2014. The algorithmic anatomy of model-based evaluation. Philos. Trans. R. Soc. B Biol. Sci. 369, 20130478. https://doi.org/10.1098/rstb.2013.0478

Daw, N.D., Niv, Y., Dayan, P., 2005. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat. Neurosci. 8, 1704–1711. https://doi.org/10.1038/nn1560

Delgado, M.R., Frank, R.H., Phelps, E.A., 2005. Perceptions of moral character modulate the neural systems of reward during the trust game. Nat. Neurosci. 8, 1611–1618. https://doi.org/10.1038/nn1575

Devaine, M., Hollard, G., Daunizeau, J., 2014. The Social Bayesian Brain: Does Mentalizing Make a Difference When We Learn? PLoS Comput. Biol. 10. https://doi.org/10.1371/journal.pcbi.1003992

Devaine, M., San-Galli, A., Trapanese, C., Bardino, G., Hano, C., Saint Jalme, M., Bouret, S., Masi, S., Daunizeau, J., 2017. Reading wild minds: A computational assay of Theory of Mind sophistication across seven primate species. PLoS Comput. Biol. 13, 1–24. https://doi.org/10.1371/journal.pcbi.1005833

Devaine, M., & Daunizeau, J., 2017. Learning about and from others' prudence, impatience or laziness: The computational bases of attitude alignment. *PLoS Computational Biology*, *13*(3), 1–28. https://doi.org/10.1371/journal.pcbi.1005422

Diaconescu, A.O., Mathys, C., Weber, L. a E., Daunizeau, J., Kasper, L., Lomakina, E.I., Fehr, E., Stephan, K.E., 2014. Inferring on the Intentions of Others by Hierarchical Bayesian Learning. PLoS Comput. Biol. 10, e1003810. https://doi.org/10.1371/journal.pcbi.1003810

Doshi, P., Gmytrasiewicz, P., 2004. A Particle Filtering Algorithm for Interactive POMDPs, in: In Proceedings of Conference on Modeling Other Agents from Observation. pp. 87–93.

Doshi, P., Qu, X., Goodie, A., 2014. Decision-Theoretic Planning in Multiagent Settings with Application to Behavioral Modeling. Plan, Act. Intent Recognit. Theory Pract. 205–224. https://doi.org/10.1016/B978-0-12-398532-3.00008-7

Doshi, P., Qu, X., Goodie, A.S., Young, D.L., 2012. Modeling human recursive reasoning using empirically informed interactive partially observable markov decision processes. IEEE Trans. Syst. Man, Cybern. Part ASystems Humans 42, 1529–1542. https://doi.org/10.1109/TSMCA.2012.2199484

Dufour, N., Redcay, E., Young, L., Mavros, P.L., Moran, J.M., Triantafyllou, C., Gabrieli, J.D.E., Saxe, R., 2013. Similar Brain Activation during False Belief Tasks in a Large Sample of Adults with and without Autism 8. https://doi.org/10.1371/journal.pone.0075468

Dumas, G., Nadel, J., Soussignan, R., Martinerie, J., Garnero, L., 2010. Inter-Brain Synchronization during Social Interaction. PLoS One 5. https://doi.org/10.1371/journal.pone.0012166

Dunbar, R.I.M., 2009. The social brain hypothesis and its implications for social evolution. Ann. Hum. Biol. 36, 562–572. https://doi.org/10.1080/03014460902960289

Emery, N.J., Clayton, N.S., 2009. Comparative Social Cognition. Annu. Rev. Psychol. 60, 87–113. https://doi.org/10.1146/annurev.psych.60.110707.163526

Emonds, G., Declerck, C.H., Boone, C., Vandervliet, E.J.M., Parizel, P.M., 2012. The cognitive demands on cooperation in social dilemmas: An fMRI study. Soc. Neurosci. 7, 494–509. https://doi.org/10.1080/17470919.2012.655426

Engel, C., Zhurakhovska, L., 2016. When is the risk of cooperation worth taking? The prisoner’s dilemma as a game of multiple motives. Appl. Econ. Lett. 23, 1157–1161. https://doi.org/10.1080/13504851.2016.1139672

Fischbacher, U., Gächter, S., 2010. Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Goods Experiments. Am. Econ. Rev. 100, 541–556. https://doi.org/10.1257/aer.100.1.541

Fishman, I., Keown, C.L., Lincoln, A.J., Pineda, J.A., Müller, R.-A., 2014. Atypical Cross Talk Between Mentalizing and Mirror Neuron Networks in Autism Spectrum Disorder. JAMA Psychiatry 71, 751. https://doi.org/10.1001/jamapsychiatry.2014.83

Frith, C.D., Frith, U., 2006. The Neural Basis of Mentalizing. Neuron 50, 531–534. https://doi.org/10.1016/j.neuron.2006.05.001

Gallagher, H.L., Frith, C.D., 2003. Functional imaging of ‘theory of mind.’ Trends Cogn. Sci. 7, 77–83. https://doi.org/10.1016/S1364-6613(02)00025-6

Garrison, J., Erdeniz, B., Done, J., 2013. Prediction error in reinforcement learning: A meta-analysis of neuroimaging studies. Neurosci. Biobehav. Rev. 37, 1297–1310. https://doi.org/10.1016/j.neubiorev.2013.03.023

Gernsbacher, M.A., Yergeau, M., 2019. Empirical failures of the claim that autistic people lack a theory of mind. Arch. Sci. Psychol. 7, 102–118. https://doi.org/10.1037/arc0000067

Gibbons, R., 1992. Game Theory for Applied Economists. Princeton University Press. https://doi.org/10.2307/j.ctvcmxrzd

Gläscher, J., Daw, N., Dayan, P., O’Doherty, J.P., 2010. States versus Rewards: Dissociable Neural Prediction Error Signals Underlying Model-Based and Model-Free Reinforcement Learning. Neuron 66, 585–595. https://doi.org/10.1016/j.neuron.2010.04.016

Gläscher, J.P., O’Doherty, J.P., 2010. Model□based approaches to neuroimaging: combining reinforcement learning theory with fMRI data. Wiley Interdiscip. Rev. Cogn. Sci. 1, 501–510. https://doi.org/10.1002/wcs.57

Gmytrasiewicz, P.J., Doshi, P., 2005. A Framework for Sequential Planning in Multi-Agent Settings. J. Artif. Intell. Res. 24, 49–79. https://doi.org/10.1613/jair.1579

Hall, K., Oram, M.W., Campbell, M.W., Eppley, T.M., Byrne, R.W., de Waal, F.B.M., 2017. Chimpanzee uses manipulative gaze cues to conceal and reveal information to foraging competitor. Am. J. Primatol. 79, e22622. https://doi.org/10.1002/ajp.22622

Hampton, A.N., Bossaerts, P., O’Doherty, J.P., 2008. Neural correlates of mentalizing-related computations during strategic interactions in humans. Proc. Natl. Acad. Sci. U. S. A. 105, 1–6. https://doi.org/10.1073/pnas.0711099105

Hedden, T., Zhang, J., 2002. What do you think I think you think?: Strategic reasoning in matrix games. Cognition 85, 1–36. https://doi.org/10.1016/S0010-0277(02)00054-9

Herbet, G., Lafargue, G., Bonnetblanc, F., Moritz-Gasser, S., Menjot de Champfleur, N., Duffau, H., 2014. Inferring a dual-stream model of mentalizing from associative white matter fibres disconnection. Brain 137, 944–959. https://doi.org/10.1093/brain/awt370

Hill, C.A., Suzuki, S., Polania, R., Moisa, M., O’Doherty, J.P., Ruff, C.C., 2017. A causal account of the brain network computations underlying strategic social behavior. Nat. Neurosci. 20, 1142–1149. https://doi.org/10.1038/nn.4602

Hill, M.R., Boorman, E.D., Fried, I., 2016. Observational learning computations in neurons of the human anterior cingulate cortex. Nat. Commun. 7, 12722. https://doi.org/10.1038/ncomms12722

Ho, T.-H., Su, X., 2013. A Dynamic Level- k Model in Sequential Games. Manage. Sci. 59, 452–469. https://doi.org/10.1287/mnsc.1120.1645

Holroyd, C.B., Coles, M.G.H., 2002. The neural basis of human error processing:

Reinforcement learning, dopamine, and the error-related negativity. Psychol. Rev. 109, 679–709. https://doi.org/10.1037/0033-295X.109.4.679

Houser, D., McCabe, K., 2014. Experimental Economics and Experimental Game Theory, in: Neuroeconomics. Elsevier, pp. 19–34. https://doi.org/10.1016/B978-0-12-416008-8.00002-4

Hsu, M., Bhatt, M., Adolphs, R., Tranel, D., Camerer, C.F., 2005. Neuroscience: Neural systems responding to degrees of uncertainty in human decision-making. Science (5754). 310, 1680–1683. https://doi.org/10.1126/science.1115327

Hula, A., Montague, P.R., Dayan, P., 2015. Monte Carlo Planning Method Estimates Planning Horizons during Interactive Social Exchange. PLOS Comput. Biol. 11. https://doi.org/10.1371/journal.pcbi.1004254

Hula, A., Vilares, I., Lohrenz, T., Dayan, P., Montague, P.R., 2018. A model of risk and mental state shifts during social interaction. PLoS Comput. Biol. 14, 1–20. https://doi.org/10.1371/journal.pcbi.1005935

Joiner, J., Piva, M., Turrin, C., Chang, S.W.C., 2017. Social learning through prediction error in the brain. npj Sci. Learn. 2, 8. https://doi.org/10.1038/s41539-017-0009-2

Jording, M., Hartz, A., Bente, G., Schulte-Rüther, M., Vogeley, K., 2019. Inferring interactivity from gaze patterns during triadic person-object-agent interactions. Front. Psychol. 10, 1–11. https://doi.org/10.3389/fpsyg.2019.01913

Kaelbling, L.P., Littman, M.L., Cassandra, A.R., 1998. Planning and Acting in Partially Observable Stochastic Domains. Artif. Intell. 101, 99–134. https://doi.org/10.1016/S0004-3702(98)00023-X

Kalbe, E., Schlegel, M., Sack, A.T., Nowak, D.A., Dafotakis, M., Bangard, C., Brand, M., Shamay-Tsoory, S., Onur, O.A., Kessler, J., 2010. Dissociating cognitive from affective theory of mind: A TMS study. Cortex 46, 769–780. https://doi.org/10.1016/J.CORTEX.2009.07.010

Karg, K., Schmelz, M., Call, J., Tomasello, M., 2016. Differing views: Can chimpanzees do Level 2 perspective-taking? Anim. Cogn. 19, 555–564. https://doi.org/10.1007/s10071-016-0956-7

Khalvati, K., Park, S. A., Mirbagheri, S., Philippe, R., Sestito, M., Dreher, J. C., & Rao, R. P. N., 2019. Modeling other minds: Bayesian inference explains human choices in group decision-making. *Science Advances*, *5*(11). https://doi.org/10.1126/sciadv.aax8783

King-Casas, B., Sharp, C., Lomax-Bream, L., Lohrenz, T., Fonagy, P., Read Montague, P., 2008. The rupture and repair of cooperation in borderline personality disorder. Science (5890). 321, 806–810. https://doi.org/10.1126/science.1156902

King-Casas, B., Tomlin, D., Anen, C., Camerer, C.F., Quartz, S.R., Montague, P.R., 2005. Getting to Know You: Reputation and Trust in a Two-Person Economic Exchange. Science (5718). 308, 78–83. https://doi.org/10.1126/science.1108062

Kleiman-Weiner, M., Ho, M., Austerweil, J., Littmann, M.L., Tenenbaum, J., 2016. Coordinate to cooperate or compete: Abstract goals and joint intentions in social interaction. Proc. 38th Annu. Conf. Cogn. Sci. Soc. 1, 1679–1684.

Kovacs, A.M., Teglas, E., Endress, A.D., 2010. The Social Sense: Susceptibility to Others' Beliefs in Human Infants and Adults. Science (6012). 330, 1830–1834. https://doi.org/10.1126/science.1190792

Krueger, F., McCabe, K., Moll, J., Kriegeskorte, N., Zahn, R., Strenziok, M., Heinecke, A., Grafman, J., 2007. Neural correlates of trust. Proc. Natl. Acad. Sci. 104, 20084–20089. https://doi.org/10.1073/pnas.0710103104

Krupenye, C., Kano, F., Hirata, S., Call, J., Tomasello, M., 2016. Great apes anticipate that other individuals will act according to false beliefs. Science (6308). 354, 110–114. https://doi.org/10.1126/science.aaf8110

Kumar, S., Rusch, T., Doshi, P., Spezio, M., Gläscher, J., 2019. Modeling cooperative and competitive decision-making in the Tiger Task, in: The Multi-Disciplinary Conference on Reinforcement Learning and Decision Making.

Kurokawa, S., Ihara, Y., 2009. Emergence of cooperation in public goods games. Proc. R. Soc. B Biol. Sci. 276, 1379–1384. https://doi.org/10.1098/rspb.2008.1546

Ledyard, J.O., 1994. Public Goods: A Survey of Experimenatl Research, in: Kagel, J., Roth, A. (Eds.), The Handbook of Experimental Economics. Princeton University Press, Printcton NJ.

Li, J., Xiao, E., Houser, D., Montague, P.R., 2009. Neural responses to sanction threats in two-party economic exchange. Proc. Natl. Acad. Sci. U. S. A. 106, 16835–40. https://doi.org/10.1073/pnas.0908855106

Loury, G.C., 1979. Market Structure and Innovation. Q. J. Econ. 93, 395. https://doi.org/10.2307/1883165

Martin, C.F., Bhui, R., Bossaerts, P., Matsuzawa, T., Camerer, C., 2014. Chimpanzee choice rates in competitive games match equilibrium game theory predictions. Sci. Rep. 4, 1–6. https://doi.org/10.1038/srep05182

McKelvey, R.D., Palfrey, T.R., 1992. An Experimental Study of the Centipede Game. Econometrica 60, 803. https://doi.org/10.2307/2951567

Mitchell, J.P., 2009. Inferences about mental states. Philos. Trans. R. Soc. B Biol. Sci. 364, 1309–1316. https://doi.org/10.1098/rstb.2008.0318

Mitchell, R.L.C., Phillips, L.H., 2015. Neuropsychologia The overlapping relationship between emotion perception and theory of mind. Neuropsychologia 70, 1–10. https://doi.org/10.1016/j.neuropsychologia.2015.02.018

Mookherjee, D., Sopher, B., 1994. Learning Behavior in an Experimental Matching Pennies Game. Games Econ. Behav. 7, 62–91. https://doi.org/10.1006/game.1994.1037

Moutoussis, M., Dolan, R. J., & Dayan, P., 2016. How People Use Social Information to Find out What to Want in the Paradigmatic Case of Inter-temporal Preferences. *PLoS Computational Biology*, *12*(7), 1–17. https://doi.org/10.1371/journal.pcbi.1004965

Nakajima, R., Kinoshita, M., Okita, H., Yahata, T., Matsui, M., Nakada, M., 2018a. Neural networks mediating high-level mentalizing in patients with right cerebral hemispheric gliomas. Front. Behav. Neurosci. 12, 1–12. https://doi.org/10.3389/fnbeh.2018.00033

Nakajima, R., Yordanova, Y.N., Duffau, H., Herbet, G., 2018b. Neuropsychological evidence for the crucial role of the right arcuate fasciculus in the face-based mentalizing network: A disconnection analysis. Neuropsychologia 115, 179–187. https://doi.org/10.1016/j.neuropsychologia.2018.01.024

Ng, A.Y., Russell, S.J., 2000. Algorithms for inverse reinforcement learning. Icml 1, 2.

Norris, A.E., Weger, H., Bullinger, C., Bowers, A., 2014. Quantifying engagement: Measuring player involvement in human–avatar interactions. Comput. Human Behav. 34, 1–11. https://doi.org/10.1016/j.chb.2014.01.044

Pearson, A., Ropar, D., Hamilton, A., 2013. A review of visual perspective taking in autism spectrum disorder . Front. Hum. Neurosci. .

Penn, D.C., Povinelli, D.J., 2007. On the lack of evidence that non-human animals possess anything remotely resembling a ‘theory of mind.’ Philos. Trans. R. Soc. B Biol. Sci. 362, 731–744. https://doi.org/10.1098/rstb.2006.2023

Phan, K.L., Sripada, C.S., Angstadt, M., McCabe, K., 2010. Reputation for reciprocity engages the brain reward center. Proc. Natl. Acad. Sci. 107, 13099–13104. https://doi.org/10.1073/pnas.1008137107

Premack, D., 2007. Human and animal cognition: Continuity and discontinuity. Proc. Natl. Acad. Sci. U. S. A. 104, 13861–13867. https://doi.org/10.1073/pnas.0706147104

Premack, D., Woodruff, G., 1978. Does the Chimpanzee have a theory of mind? Behav.

Brain Sci. 4, 515–526. https://doi.org/10.1016/j.celrep.2011.1011.1001.7.

Puterman, M.L., 1990. Chapter 8 Markov decision processes, in: Handbooks in Operations Research and Management Science. Elsevier, pp. 331–434. https://doi.org/10.1016/S0927-0507(05)80172-0

Qu, C., Ligneul, R., Van der Henst, J.-B., Dreher, J.-C., 2017. An Integrative Interdisciplinary Perspective on Social Dominance Hierarchies. Trends Cogn. Sci. 21, 893–908. https://doi.org/10.1016/j.tics.2017.08.004

Redcay, E., Schilbach, L., 2019. Using second-person neuroscience to elucidate the mechanisms of social interaction. Nat. Rev. Neurosci. 20, 495–505. https://doi.org/10.1038/s41583-019-0179-4

Rilling, J.K., Gutman, D.A., Zeh, T.R., Pagnoni, G., Berns, G.S., Kilts, C.D., 2002. A Neural Basis for Social Cooperation. Neuron 35, 395–405. https://doi.org/10.1016/S0896-6273(02)00755-9

Rosenthal, R.W., 1981. Games of perfect information, predatory pricing and the chain-store paradox. J. Econ. Theory 25, 92–100. https://doi.org/10.1016/0022-0531(81)90018-1

Rousseau, J.J., Cranson, M., 1984. A discourse on inequality, 6th ed. Penguin Books, Hamondsworth England.

Sanfey, A.G., 2007. Social decision-making: insights from game theory and neuroscience. Science 318, 598–602. https://doi.org/10.1126/science.1142996

Saxe, R., 2006. Uniquely human social cognition. Curr. Opin. Neurobiol. 16, 235–239. https://doi.org/10.1016/j.conb.2006.03.001

Saxe, R., Kanwisher, N., 2003. People thinking about thinking people. The role of the temporo-parietal junction in “theory of mind.” Neuroimage 19, 1835–1842. https://doi.org/10.1016/S1053-8119(03)00230-1

Schaafsma, S.M., Pfaff, D.W., Spunt, R.P., Adolphs, R., 2015. Deconstructing and reconstructing theory of mind. Trends Cogn. Sci. 19, 65–72. https://doi.org/10.1016/j.tics.2014.11.007

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., Vogeley, K., 2013. Toward a second-person neuroscience. Behav. Brain Sci. 36, 393–414. https://doi.org/10.1017/S0140525X12000660

Schmelz, M., Call, J., Tomasello, M., 2011. Chimpanzees know that others make inferences. Proc. Natl. Acad. Sci. 108, 3077–3079. https://doi.org/10.1073/pnas.1000469108

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., Perner, J., 2014. Fractionating theory of mind: A meta-analysis of functional brain imaging studies. Neurosci. Biobehav. Rev. 42, 9–34. https://doi.org/10.1016/j.neubiorev.2014.01.009

Selbing, I., Olsson, A., 2017. Beliefs about Others’ Abilities Alter Learning from Observation. Sci. Rep. 7, 16173. https://doi.org/10.1038/s41598-017-16307-3

Senju, A., Southgate, V., Miura, Y., Matsui, T., Hasegawa, T., Tojo, Y., Osanai, H., Csibra, G., 2010. Absence of spontaneous action anticipation by false belief attribution in children with autism spectrum disorder. Dev. Psychopathol. 22, 353–360. https://doi.org/10.1017/S0954579410000106

Senju, A., Southgate, V., White, S., Frith, U., 2009. Mindblind Eyes: An Absence of Spontaneous Theory of Mind in Asperger Syndrome. Science (5942). 325, 883–885. https://doi.org/10.1126/science.1176170

Shum, M., Kleiman-Weiner, M., Littman, M.L., Tenenbaum, J.B., 2019. Theory of Minds: Understanding Behavior in Groups Through Inverse Planning.

Siegal, M., Varley, R., 2002. Neural systems involved in “theory of mind.” Nat. Rev. Neurosci. 3, 463–471. https://doi.org/10.1038/nrn844

Singer, T., Kiebel, S.J., Winston, J.S., Dolan, R.J., Frith, C.D., 2004. Brain Responses to the Acquired Moral Status of Faces. Neuron 41, 653–662. https://doi.org/10.1016/S0896-

6273(04)00014-5

Skyrms, B., 2001. The Stag Hunt. Proc. Addresses Am. Philos. Assoc. 75, 31. https://doi.org/10.2307/3218711

Spezio, M.L., Huang, P.Y.S., Castelli, F., Adolphs, R., 2007. Amygdala damage impairs eye contact during conversations with real people. J. Neurosci. 27, 3994–3997. https://doi.org/10.1523/JNEUROSCI.3789-06.2007

Stahl, D.O., 1993. Evolution of Smartn Players. Games Econ. Behav. 5, 604–617. https://doi.org/10.1006/game.1993.1033

Stolk, A., Verhagen, L., Toni, I., 2015. Conceptual Alignment: How Brains Achieve Mutual Understanding. Trends Cogn. Sci. xx, 1–12. https://doi.org/10.1016/j.tics.2015.11.007

Surian, L., Caldi, S., Sperber, D., 2007. Attribution of Beliefs by 13-Month-Old Infants. Psychol. Sci. 18, 580–586. https://doi.org/10.1111/j.1467-9280.2007.01943.x

Sutton, R.S., Barto, A.G., 2012. Reinforcement Learning : An Introduction, 2nd ed. MIT Press, Cambridge, MA, USA.

Touhey, J.C., 1974. Decision Processes, Expectations, and Adoption of Strategies in Zero-Sum Games. Hum. Relations 27, 813–824. https://doi.org/10.1177/001872677402700807

Van Overwalle, F., Baetens, K., 2009. Understanding others' actions and goals by mirror and mentalizing systems: A meta-analysis. Neuroimage 48, 564–584. https://doi.org/10.1016/J.NEUROIMAGE.2009.06.009

Vivanti, G., Rogers, S.J., 2011. Action understanding and social learning in autism: A developmental perspective. Life Span Disabil. 14, 7–29.

Wan Lee, S., Shimojo, S., O'Doherty, J.P., 2014. Neural Computations Underlying Arbitration between Model-Based and Model-free Learning. Neuron 81, 687–699. https://doi.org/10.1016/j.neuron.2013.11.028

Whiten, A., Byrne, R.W., 1988. The Machiavellian intelligence hypotheses: Editorial., in: Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans. Clarendon Press/Oxford University Press, New York, NY, US, pp. 1–9.

Wimmer, H., Perner, J., 1983. Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. Cognition 13, 103–128. https://doi.org/10.1016/0010-0277(83)90004-5

Xiao, Y., Geng, F., Riggins, T., Chen, G., Redcay, E., 2019. Neural correlates of developing theory of mind competence in early childhood. Neuroimage 184, 707–716. https://doi.org/10.1016/j.neuroimage.2018.09.079

Yamakawa, T., Okano, Y., Saijo, T., 2016. Detecting motives for cooperation in public goods experiments. Exp. Econ. 19, 500–512. https://doi.org/10.1007/s10683-015-9451-2

Yoshida, W., Dolan, R.J., Friston, K.J., 2008. Game Theory of Mind. PLoS Comput. Biol. 4. https://doi.org/10.1371/journal.pcbi.1000254

Yoshida, W., Dziobek, I., Kliemann, D., Heekeren, H.R., Friston, K.J., Dolan, R.J., 2010. Cooperation and heterogeneity of the autistic mind. J. Neurosci. 30, 8815–8818. https://doi.org/10.1523/JNEUROSCI.0400-10.2010

Yoshida, W., Seymour, B., Friston, K.J., Dolan, R.J., 2010. Neural Mechanisms of Belief Inference during Cooperative Games. J. Neurosci. 30, 10744–10751. https://doi.org/https://doi.org/10.1523/JNEUROSCI.5895-09.2010

Zhang, L., Gläscher, J.P., 2019. A network supporting social influences in human decision-making. bioRxiv 551614. https://doi.org/https://doi.org/10.1101/551614

Zhu, L., Jiang, Y., Scabini, D., Knight, R.T., Hsu, M., 2019. Patients with basal ganglia damage show preserved learning in an economic game. Nat. Commun. 10, 802. https://doi.org/10.1038/s41467-019-08766-1

Zhu, L., Mathewson, K.E., Hsu, M., 2011. Dissociable neural representations of reinforcement and belief prediction errors underlie strategic learning 1–6. https://doi.org/10.1073/pnas.1116783109

Highlights

- The ability to form a Theory of Mind (ToM) constitutes a hallmark of human cognition.
- We review various decision tasks and computational models aimed at ToM.
- Tasks and models are characterized with respect to interactivity & uncertainty.
- We suggest that the complexity of ToM varies along these two primary dimensions.